Internet Vision

Lecture 14

With many slides from: N. Snavely, L. van Ahn, J. Hays, A. Efros

What is Internet Vision?

- Vast majority of data on Internet is in form of images/video
- Lots of unique applications of Computer Vision in this setting
- Also a very useful tool for vision researchers
  - Get labels for images

The Internet as source of labor

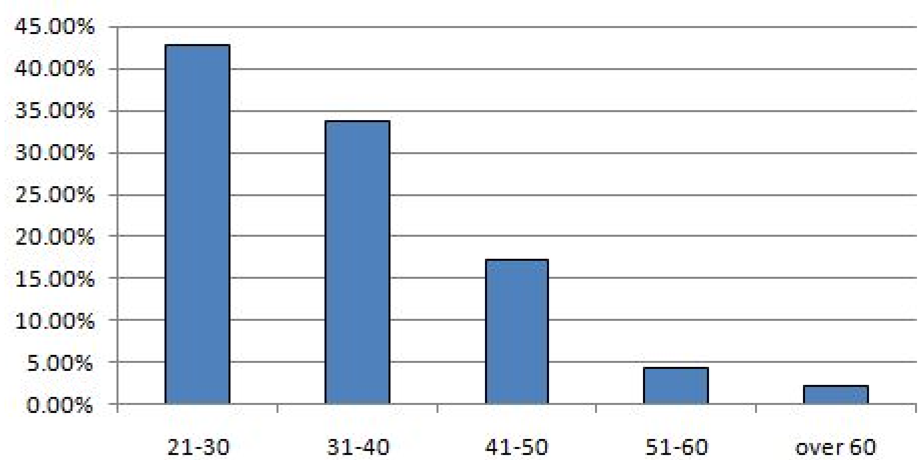
Slide credit: N. Snavely

LabelMe

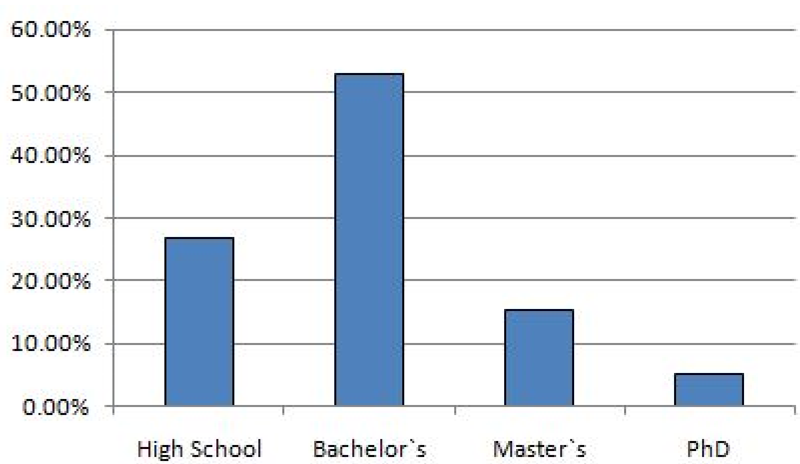

Zoom
Erase
Help
Make 3D
Upload image
Show me another image
Sign in (why?)
There are 773284 labelled objects
Polygons in this image
(IMG, XML)
tree
tree
house
sidewalk
bycicle
Sign
top of tree
ma
car
road
sky
sign
sign
door
car

Mechanical Turk – Demographics

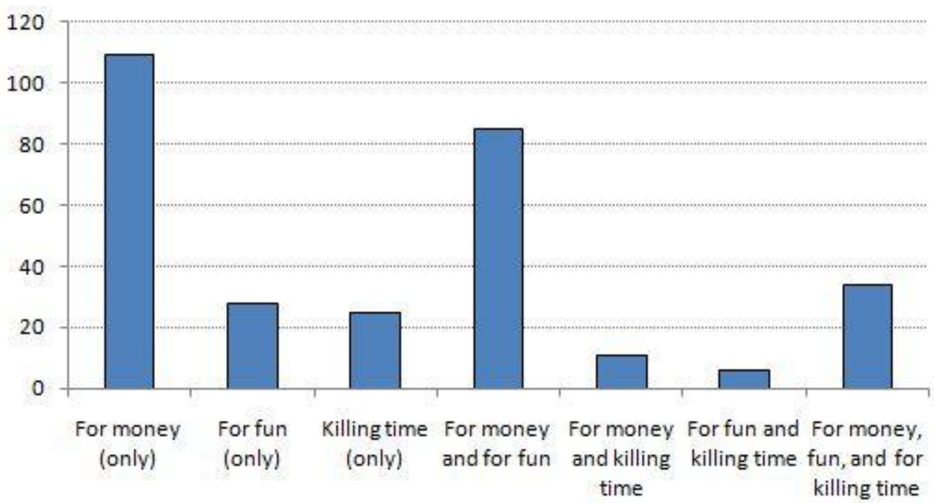
United States	76.25%
India	8.03%
United Kingdom	3.34%
Canada	2.34%



Age distribution



Age distribution



Motivation

LABELING IMAGES WITH WORDS

→

MARTHA STEWART

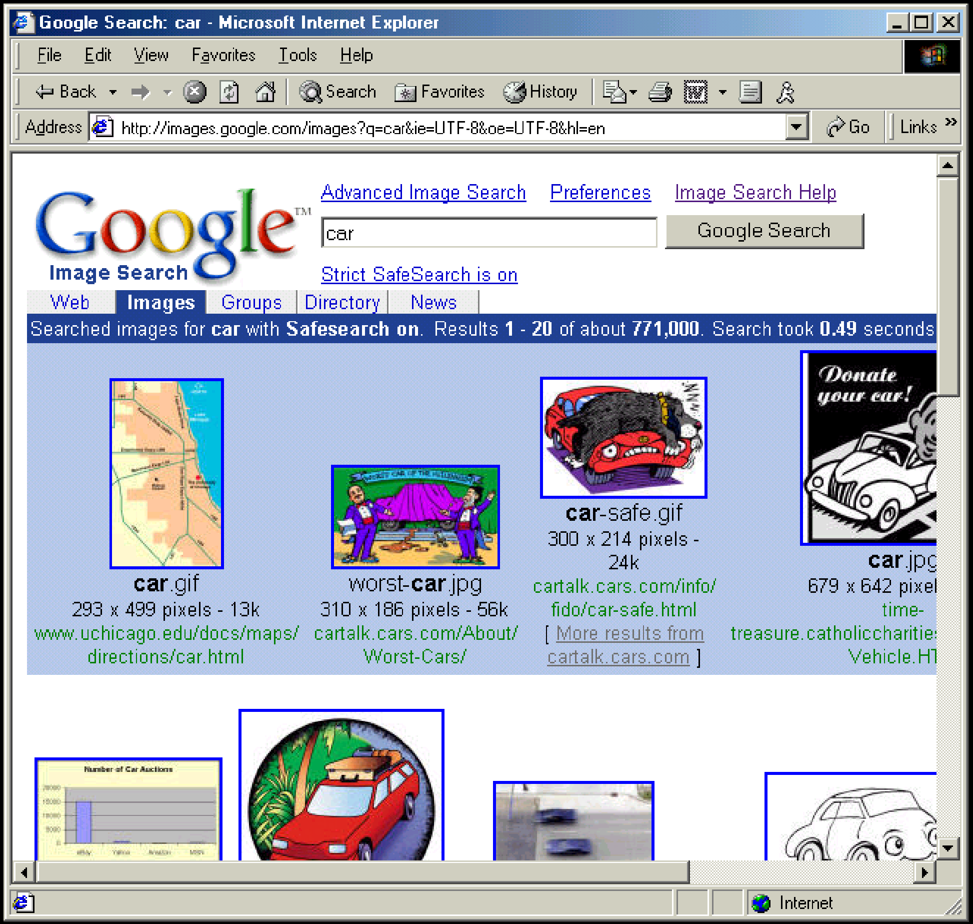
FLOWERS

SUPER EVIL

STILL AN OPEN PROBLEM

Slides courtesy Luis von Ahn

IMAGE SEARCH ON THE WEB



USES FILENAMES AND HTML TEXT

THE **ESP GAME**

TWO-PLAYER ONLINE GAME

PARTNERS DON’T KNOW EACH OTHER AND CAN’T COMMUNICATE

OBJECT OF THE GAME:
TYPE THE SAME WORD

THE ONLY THING IN COMMON IS
AN IMAGE

THE **ESP GAME**

PLAYER 1

GUESSING: CAR

GUESSING: HAT

GUESSING: KID

SUCCESS!
YOU AGREE ON CAR

PLAYER 2

GUESSING: BOY

GUESSING: CAR

SUCCESS!
YOU AGREE ON CAR


The ESP Game - Microsoft Internet Explorer
0:46
Time Left
The ESP Game
0220
score
Taboo Words
HAT
SUNGLASSES
Your Guesses
MAN
PERSON
GUY
Type your next guess:
Pass
© 2004 Carnegie Mellon University, all rights reserved. Patent Pending.

THE ESP GAME IS FUN

4.1 MILLION LABELS WITH 23,000 PLAYERS

THERE ARE MANY PEOPLE THAT PLAY
OVER 20 HOURS A WEEK

SAMPLE LABELS

BEACH
CHAIRS
SEA
PEOPLE
MAN
WOMAN
PLANT
OCEAN
TALKING
WATER
PORCH

REVEALING IMAGES

GUESSER

BRUSH

GUESS

REVEALER

CAR

BRUSH

PARTNER'S GUESS

Photo Collections

- Phototourism / Photosynth
  - Snavely, Szeliski and Seitz (Siggraph 2006)

Scene exploration

Mapping the World's Photos (35 million)

Slide credit: N. Snavely

[Crandall, Backstrom, Huttenlocher, Kleinberg, WWW '09]

Mapping the World's Photos

Slide credit: N. Snavely

[Crandall, Backstrom, Huttenlocher, Kleinberg, WWW '09]

Camera calibration

"Priors for Large Photo Collections and What They Reveal about Cameras ,"
S. Kuthirummal, A. Agarwala, D. B Goldman, and S. K. Nayar,
European Conference on Computer Vision, 2010

Slide credit: N. Snavely

Leveraging Huge Data

- What if we had millions or billions of images?
  - Facebook has O(10^10) images (10 Billion)
  - Roughly a lifetime of visual experience (5 glances/sec)

- What kind of new algorithms could we apply?
  - Brute Force methods

Scene Completion Using Millions of Photographs

James Hays and Alexei A. Efros

Carnegie Mellon University

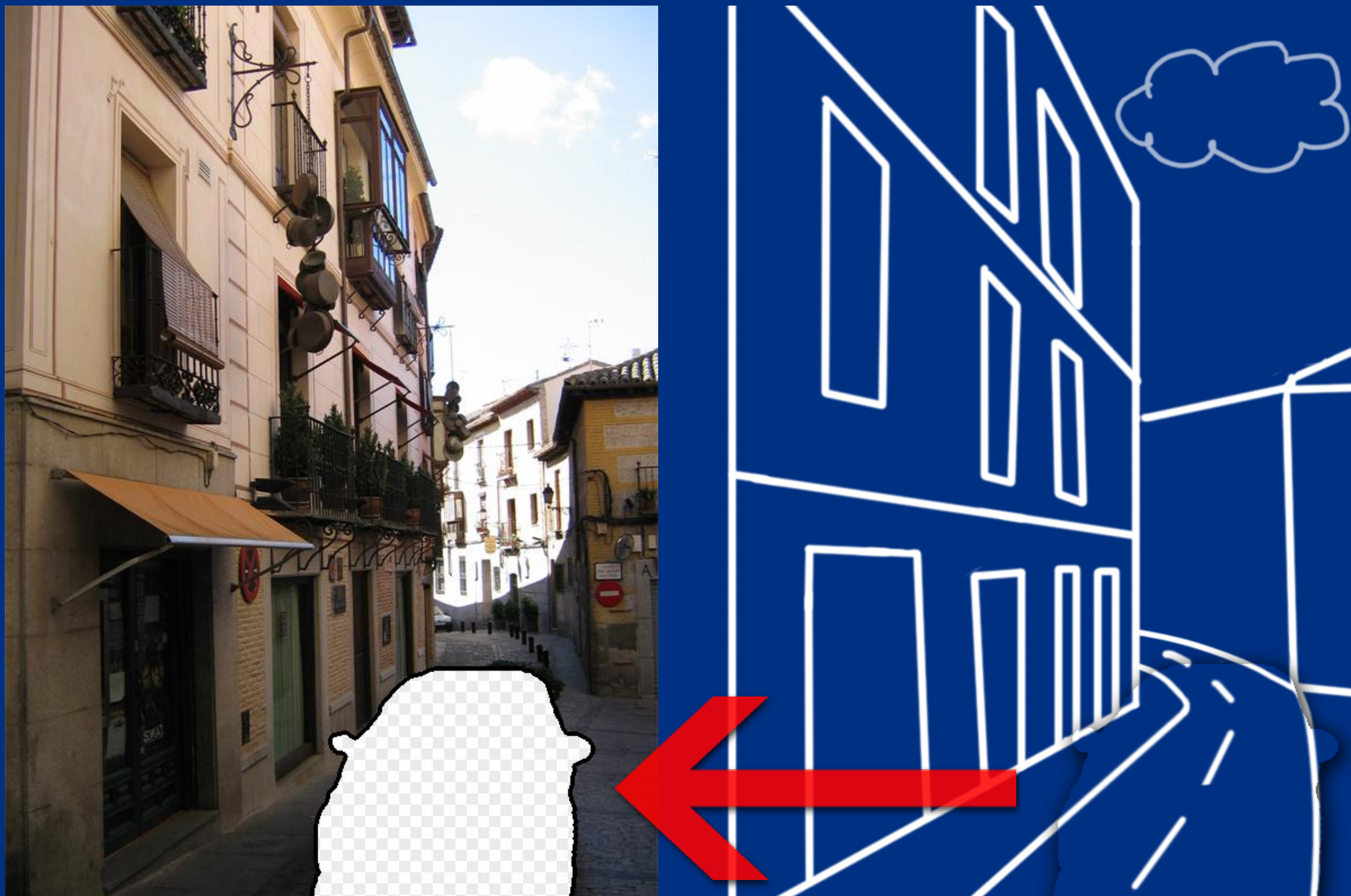

5407 BWL

Efros and Leung result

Criminisi et al. result

Criminisi et al. result

Scene Matching for Image Completion



Google

alley | Search Images | Search the Web | Advanced Image Search | Preferences

Strict SafeSearch is on

Images Showing: All image sizes

Results **1 - 20** of about **908,000** for **alley** [definition] with **Safesearch on**. (**0.07** seconds)

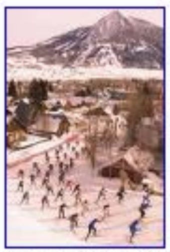
Change **Alley** Aerial Plaza with its ...
300 x 400 - 21k
en.wikipedia.org

The Printer's **Alley** sign looking ...
679 x 450 - 469k - jpg
franklin.thefuntimesguide.com

Looking west past Printers **Alley**.
679 x 450 - 464k - jpg
franklin.thefuntimesguide.com

More Bubble Gum **Alley** photos can be ...
764 x 591 - 33k - gif
www.locallinks.com

Gasoline **Alley** gang
692 x 430 - 177k - jpg
newcritics.com



2007 **Alley** Loop Sponsors
300 x 453 - 51k - jpg
www.cbnordic.org

Change **Alley** : interior
550 x 413 - 98k
infopedia.nlb.gov.sg

Earl G. **Alley** ...
321 x 383 - 19k - jpg
www.msstate.edu

Gun **Alley** 8.5x11 Full Color Ink Wash ...
390 x 301 - 14k - jpg
www.rorschachentertainment.com

Grace Court **Alley**
732 x 549 - 98k - jpg
www.bridgeandtunnelclub.com

Grace Court **Alley**
732 x 549 - 80k - jpg
www.bridgeandtunnelclub.com

panoramic photo of Alligator **Alley**
4902 x 460 - 1048k - jpg
sflwww.er.usgs.gov

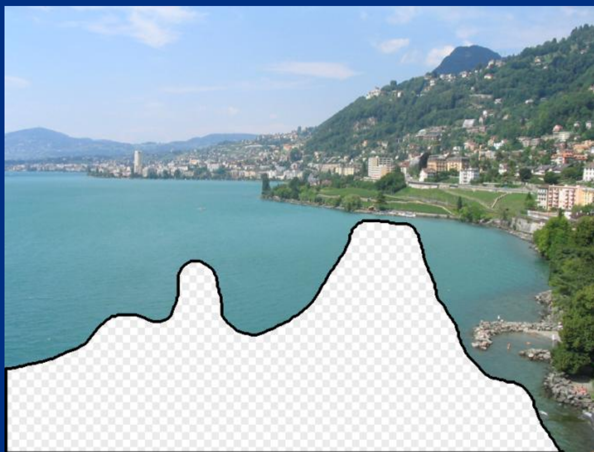
Richard B. **Alley**
450 x 361 - 29k - gif
www.ncdc.noaa.gov

Also, Chicken **Alley** is reported to ...
450 x 337 - 82k
phidoux.typepad.com

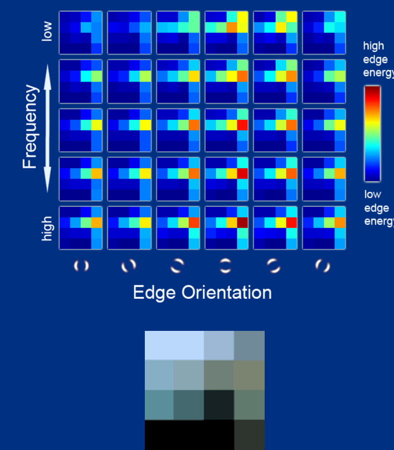
Ego **Alley**
500 x 375 - 48k - jpg
dc.about.com

Scene Completion Result

The Algorithm



Input image



Scene Descriptor

...

Image Collection

...

200 matches

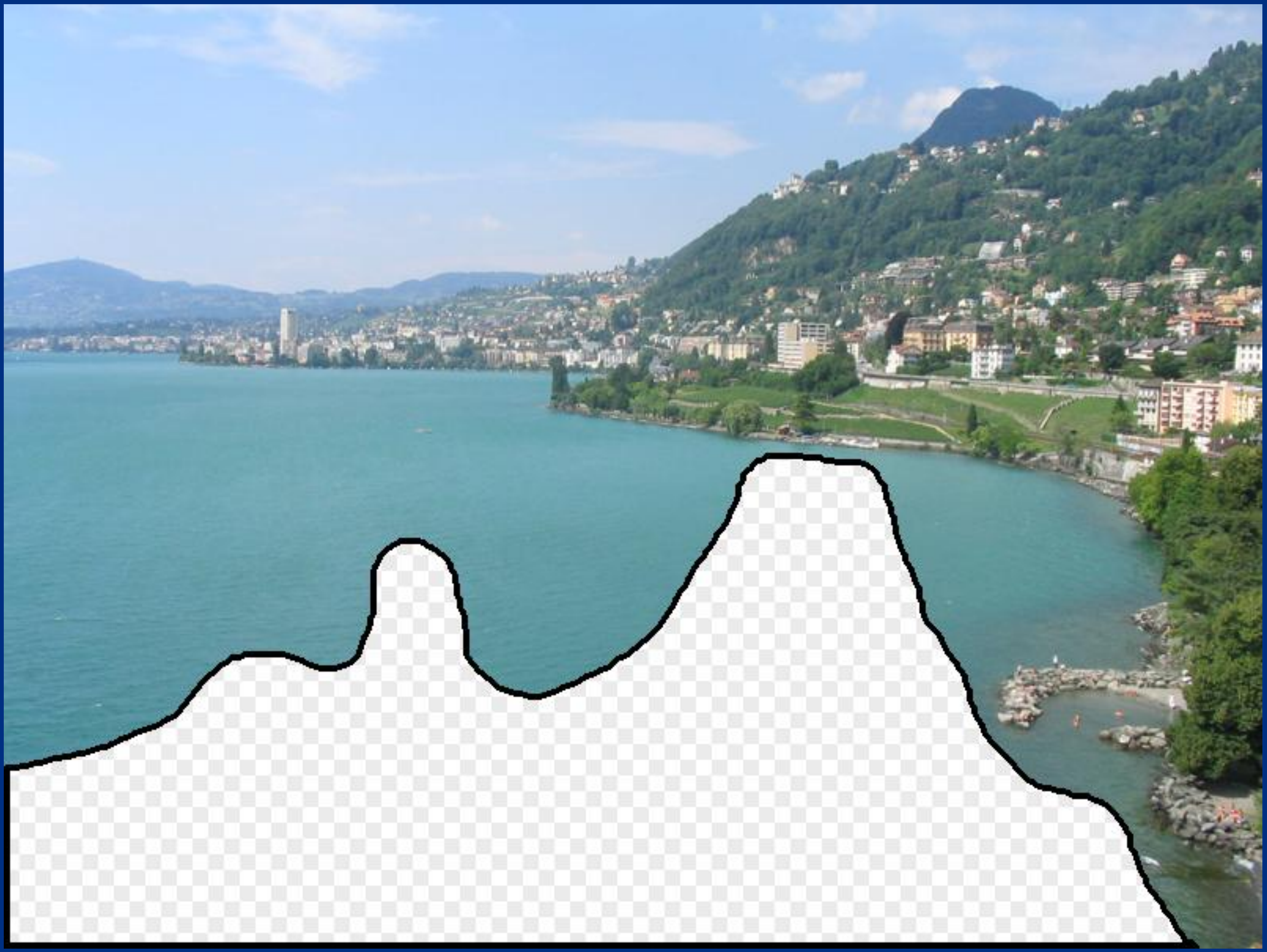
Context matching
+ blending

20 completions

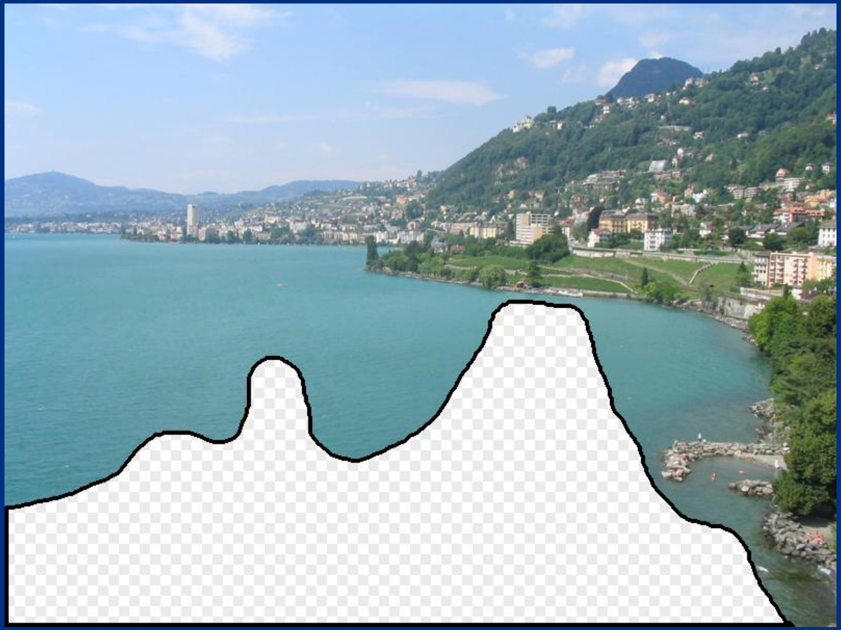
Data

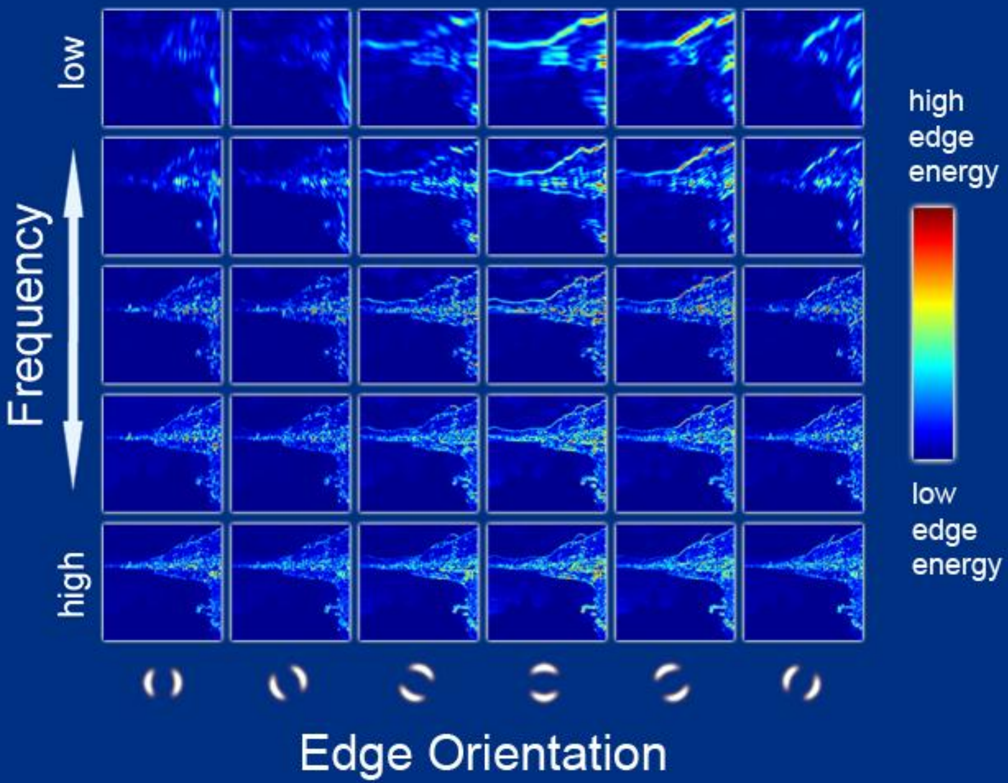
We downloaded **2.3 Million** unique images from Flickr groups and keyword searches.

Scene Matching

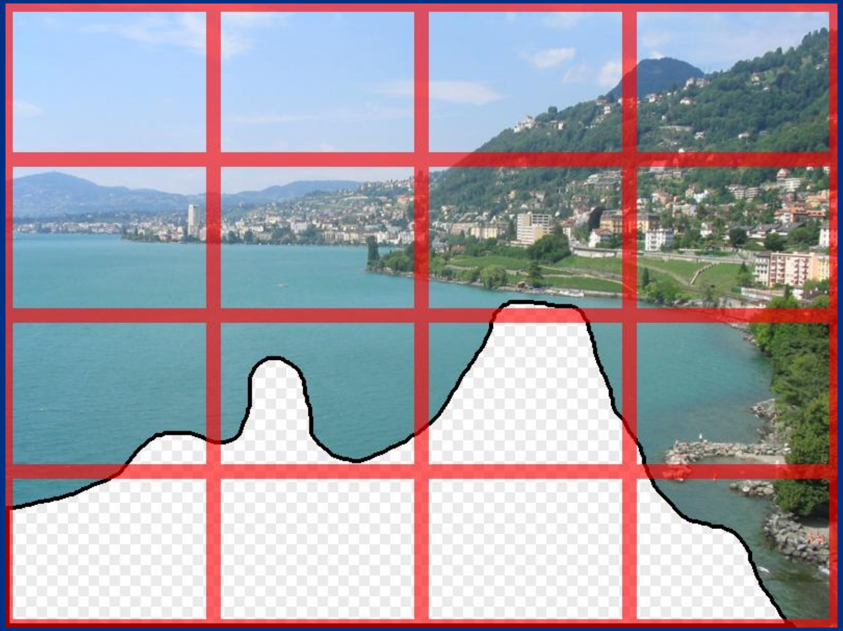


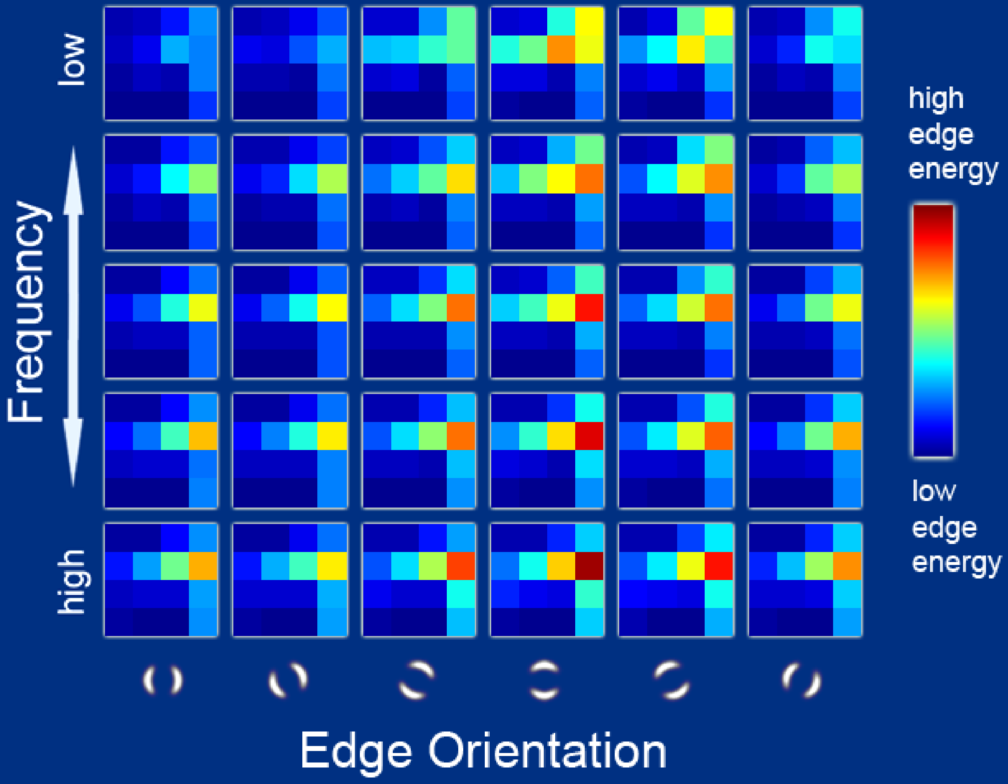
Scene Descriptor





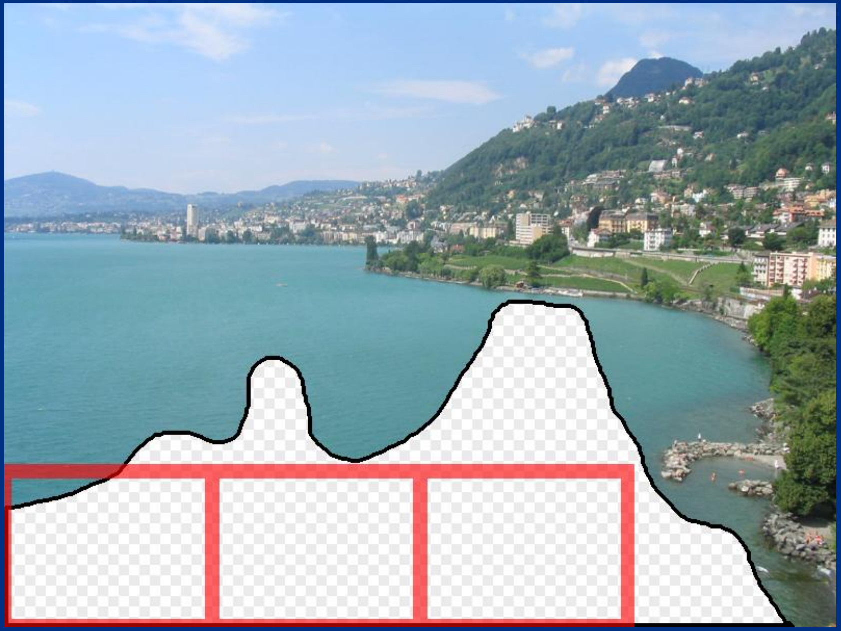
Scene Descriptor

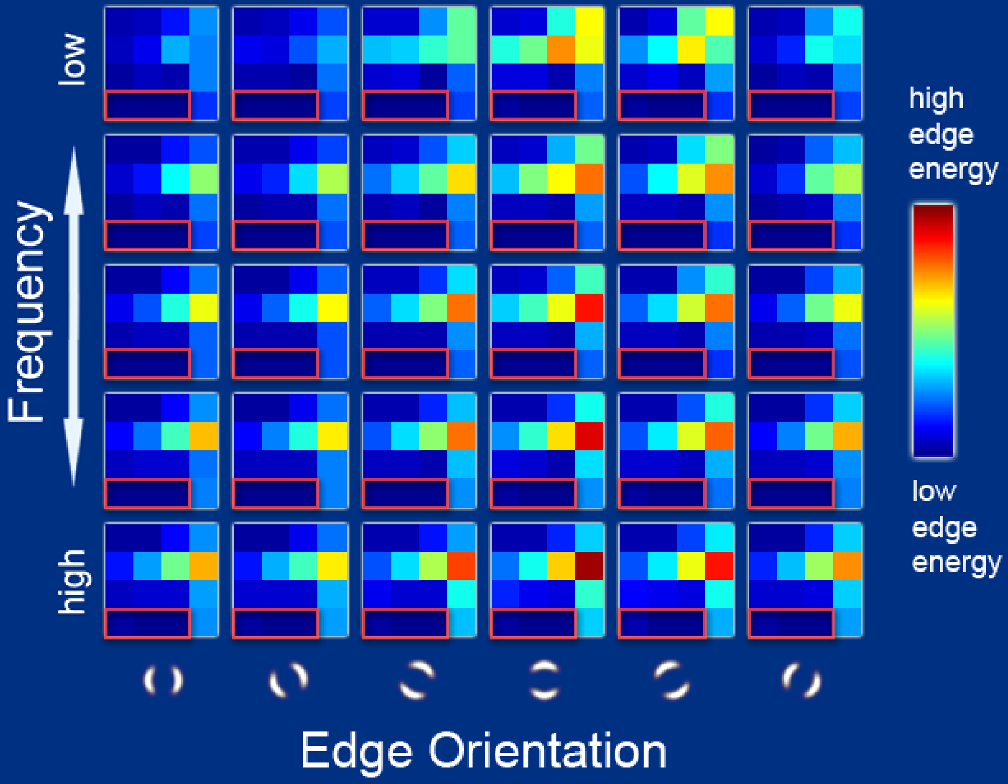




Gist scene descriptor
(Oliva and Torralba 2001)

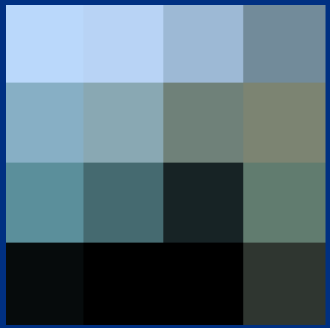
Scene Descriptor



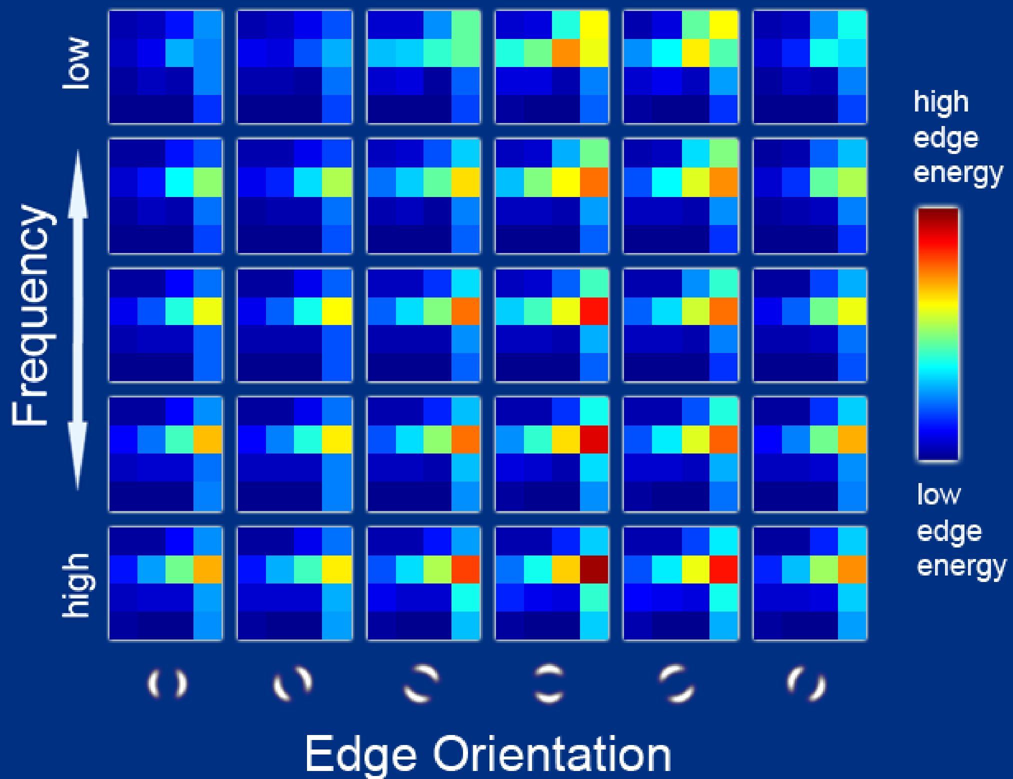


Gist scene descriptor
(Oliva and Torralba 2001)

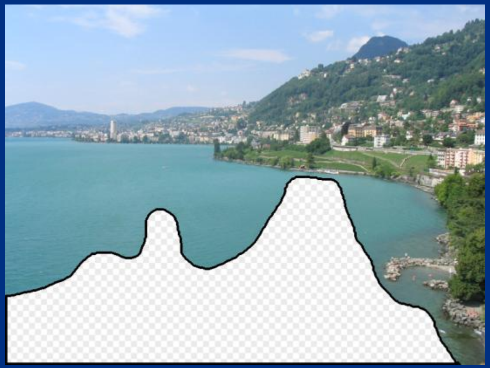
Scene Descriptor

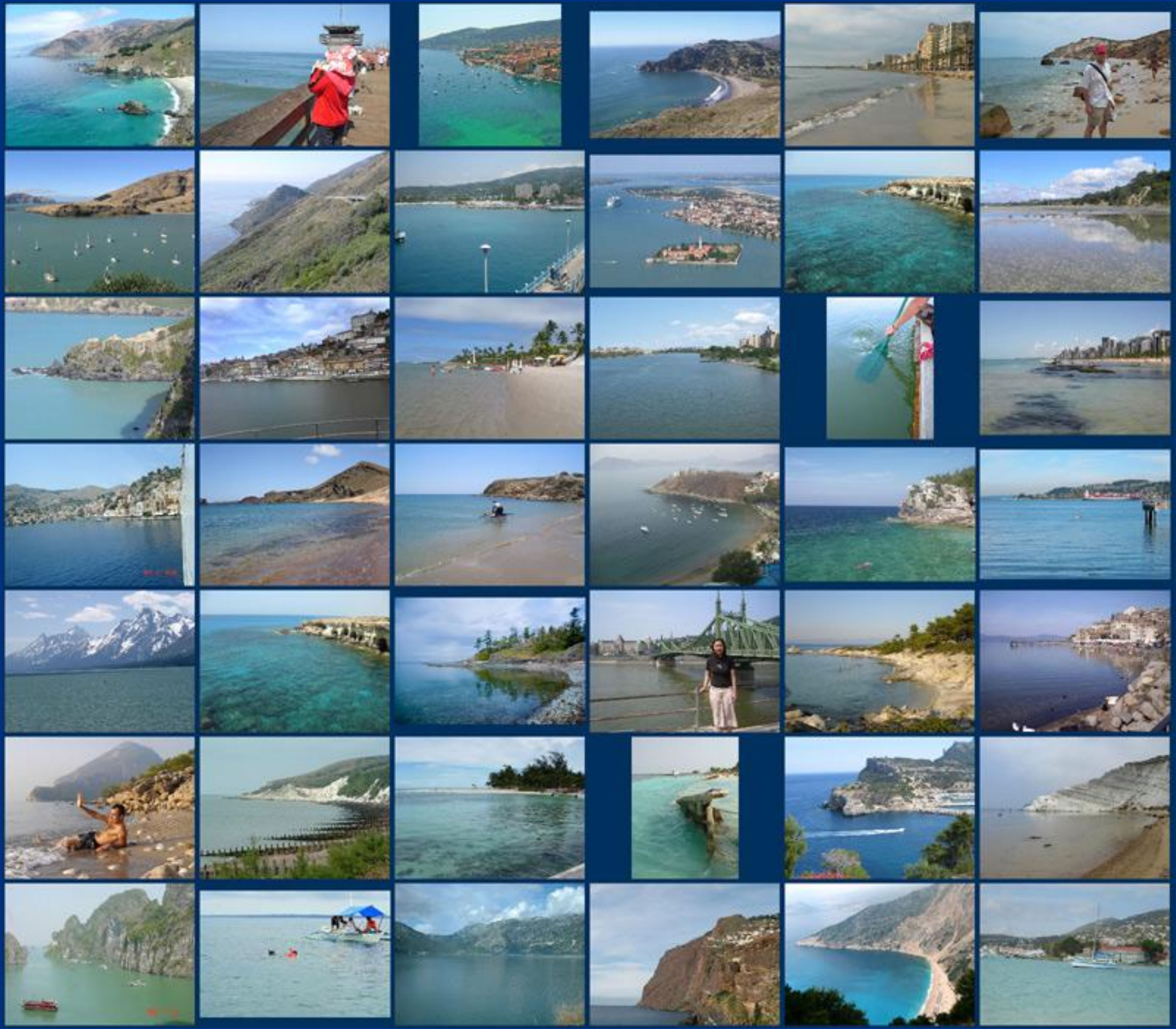


+



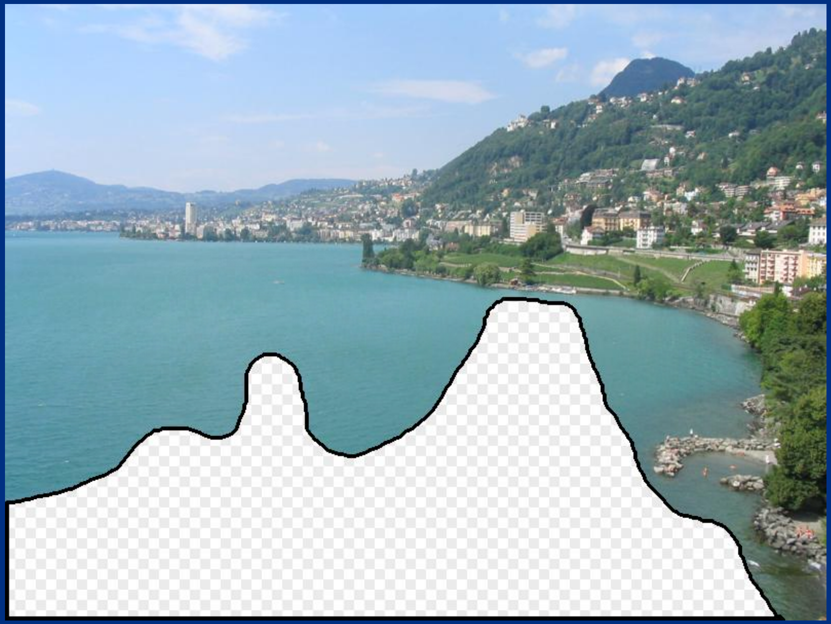
Gist scene descriptor
(Oliva and Torralba 2001)





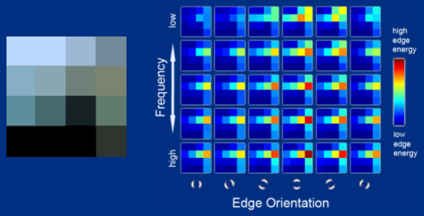
… 200 total

Context Matching



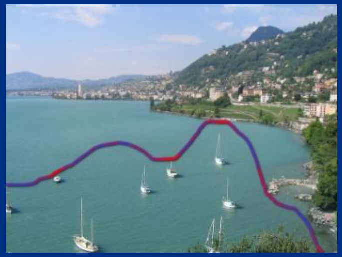
Result Ranking

We assign each of the 200 results a score which is the sum of:



The scene matching distance

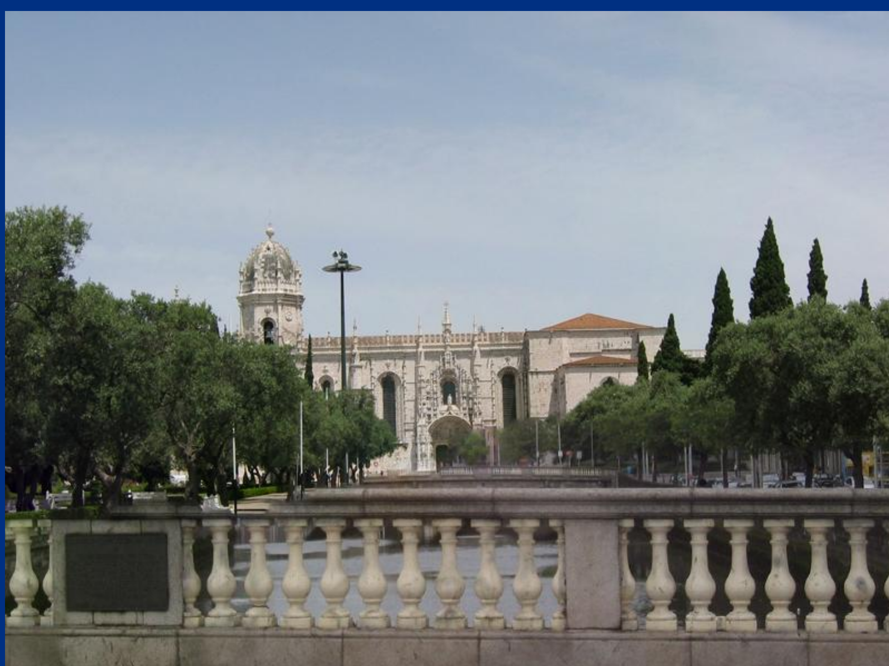
The context matching distance
(color + texture)



The graph cut cost

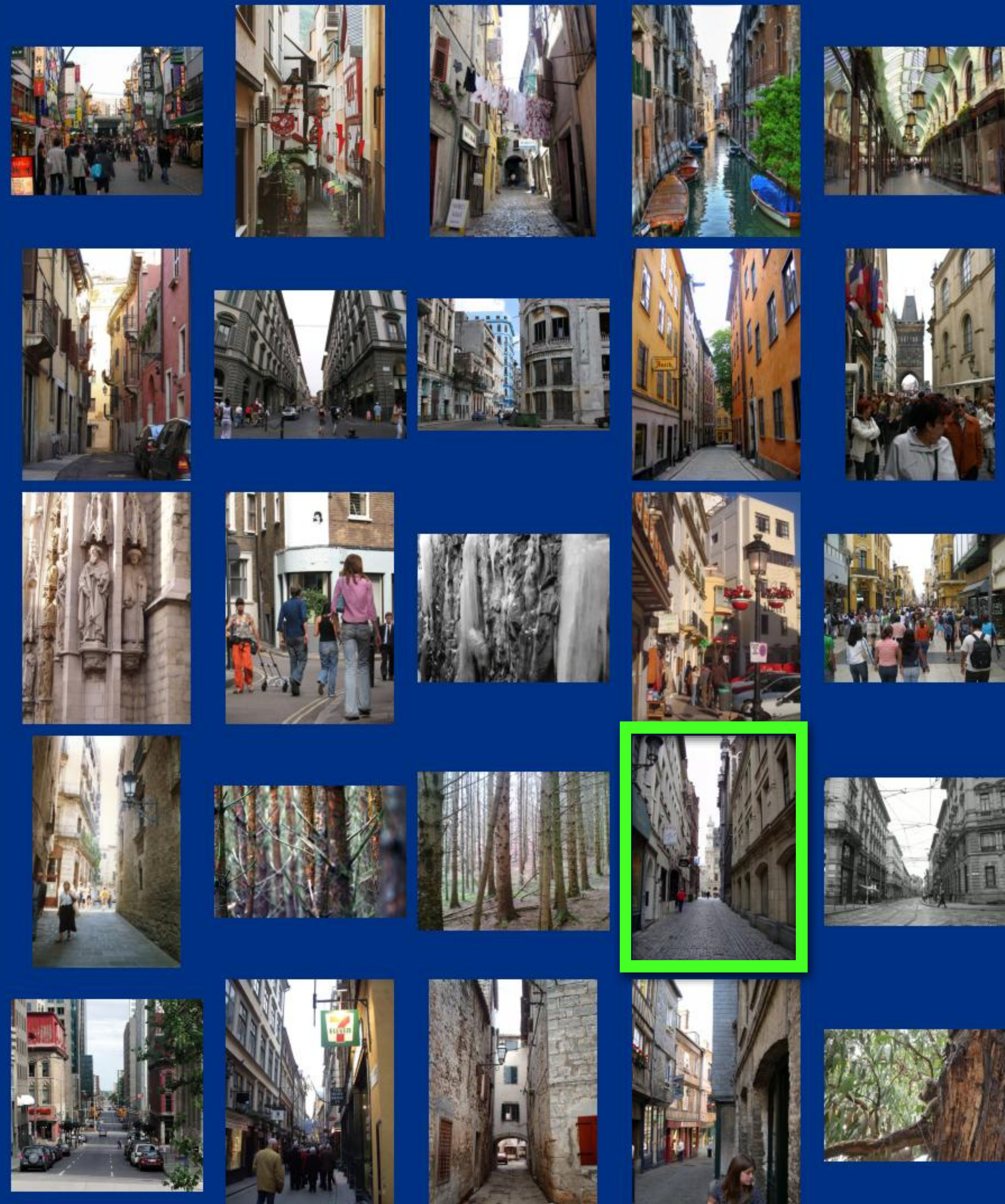
Top 20 Results


OCT 30 2003

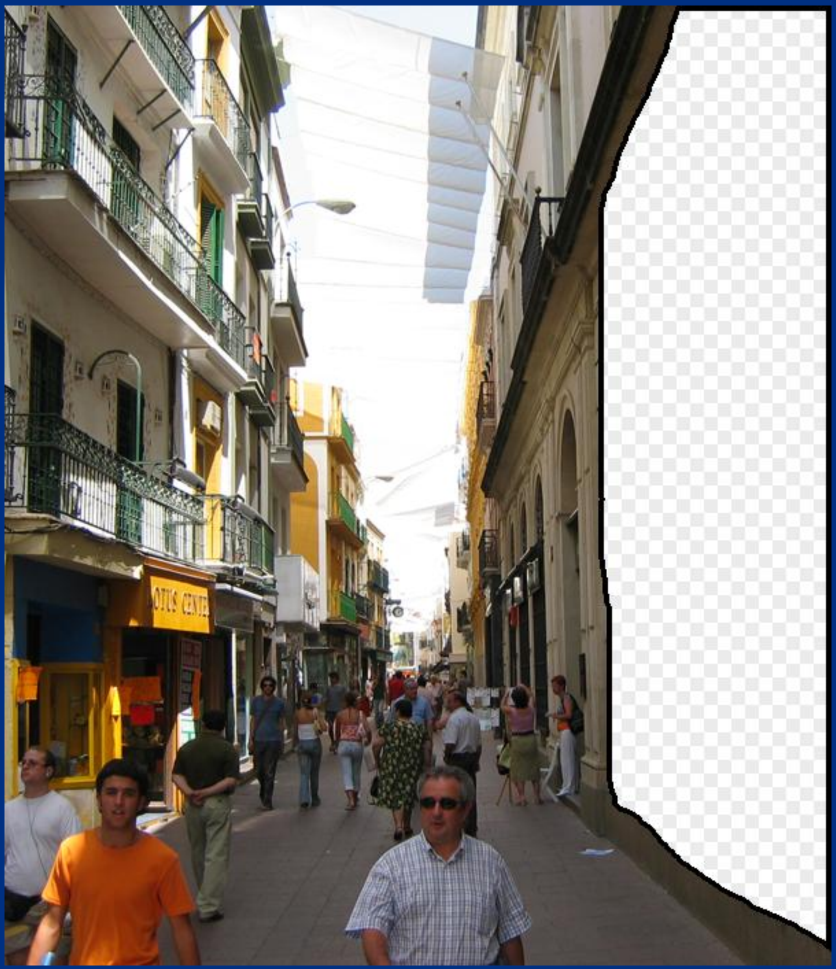



OTUS CENTE

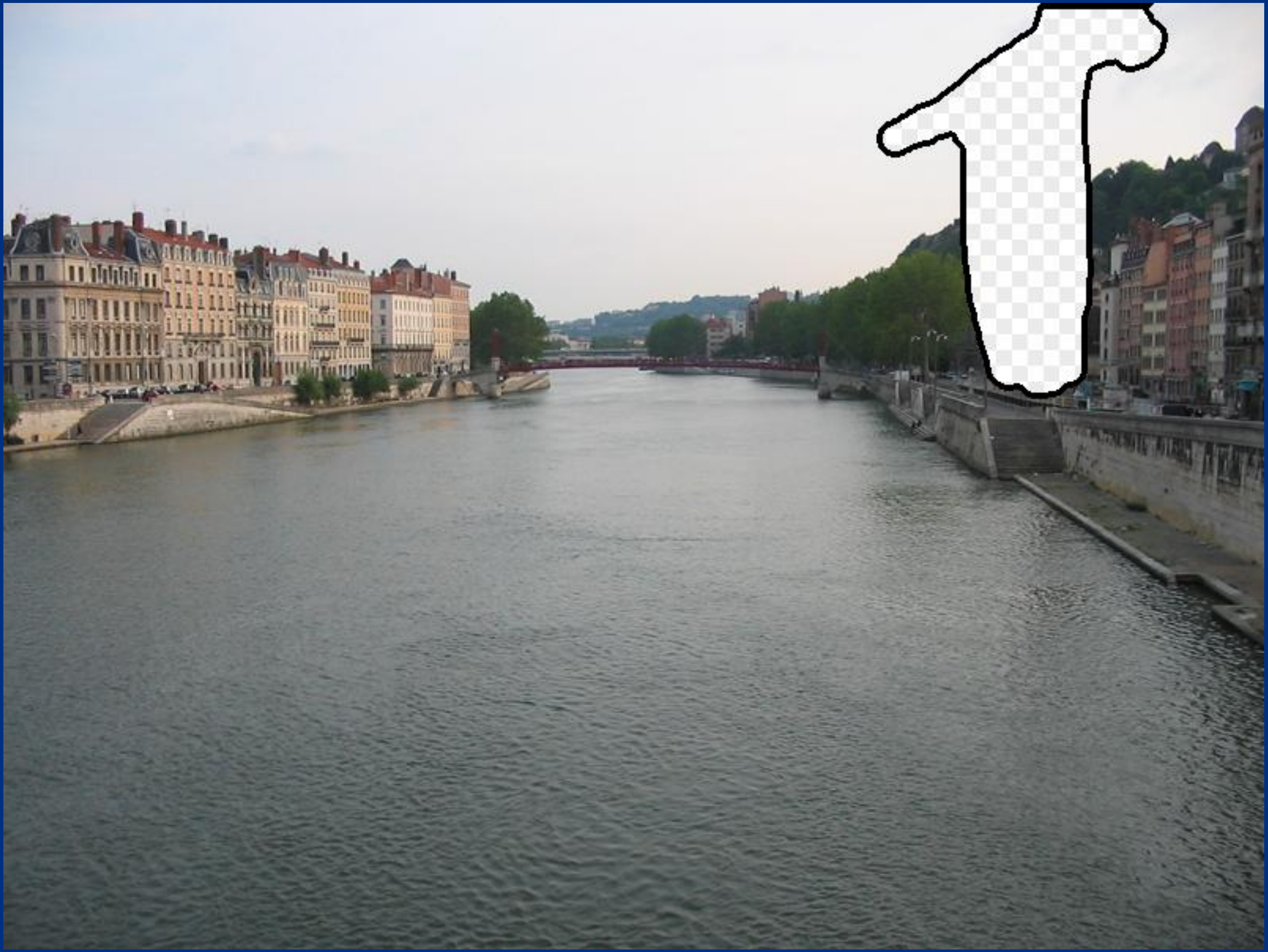
… 200 scene matches

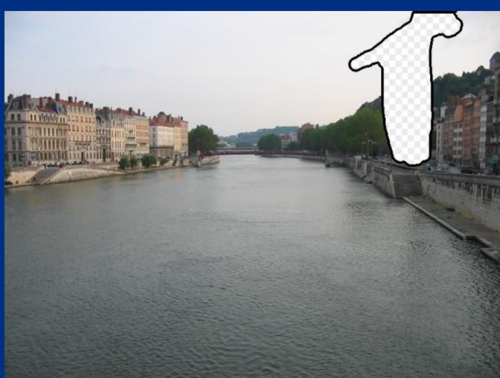


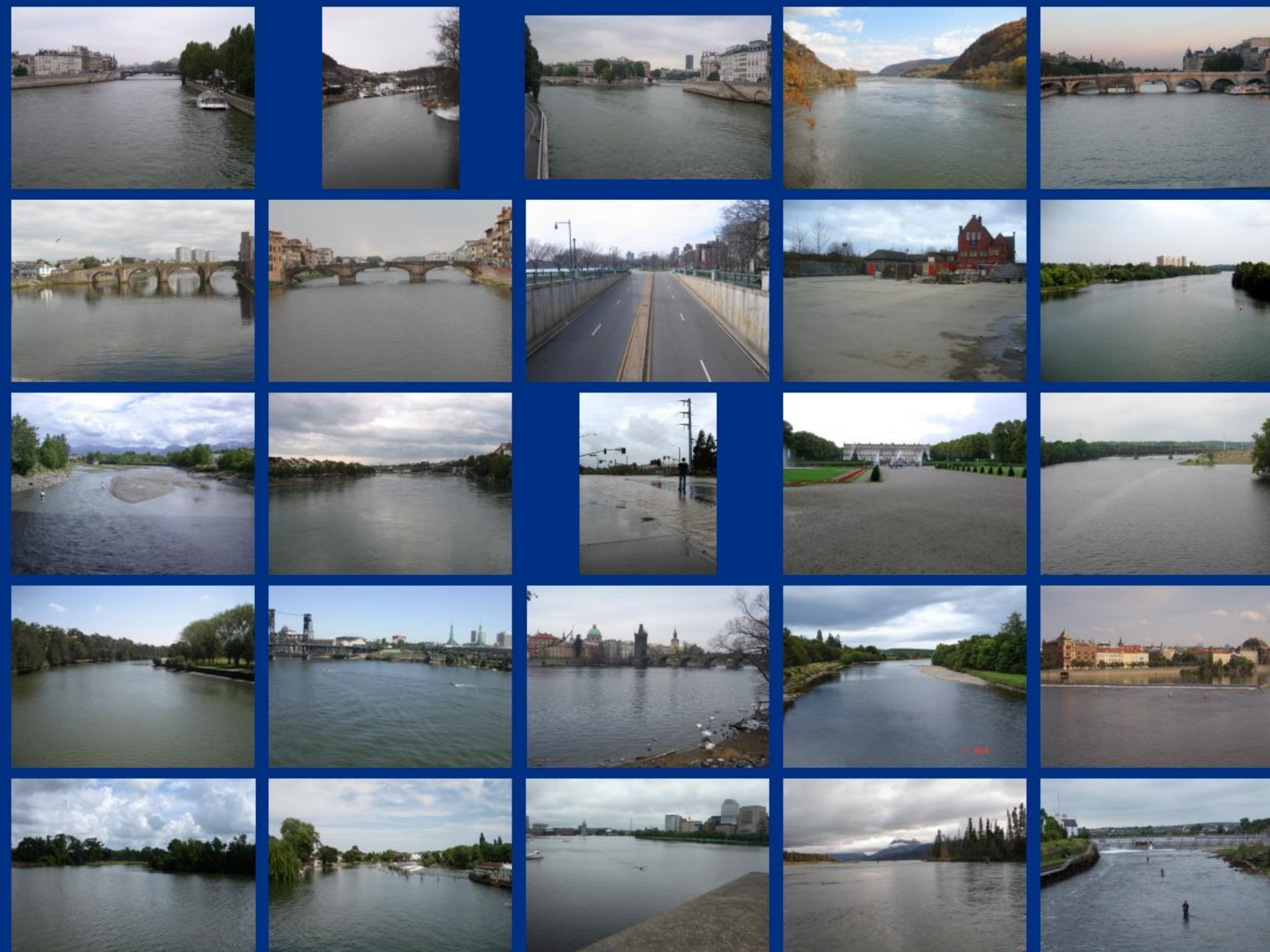
… 200 scene matches



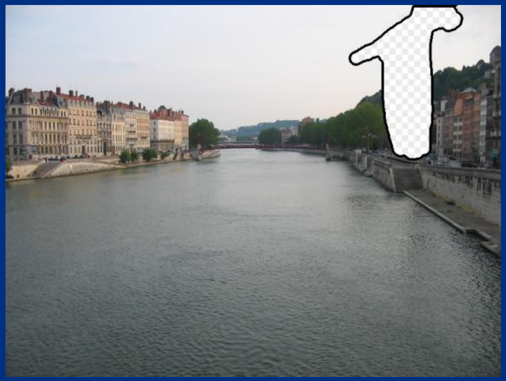

HOTEL
Jupiter



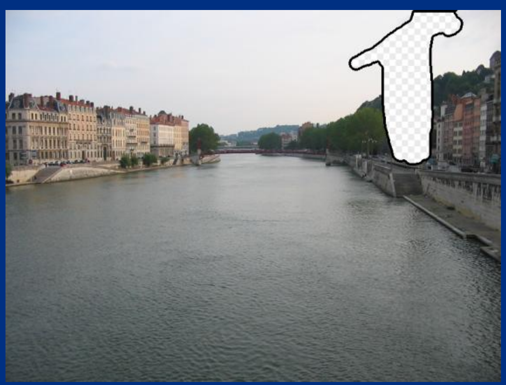




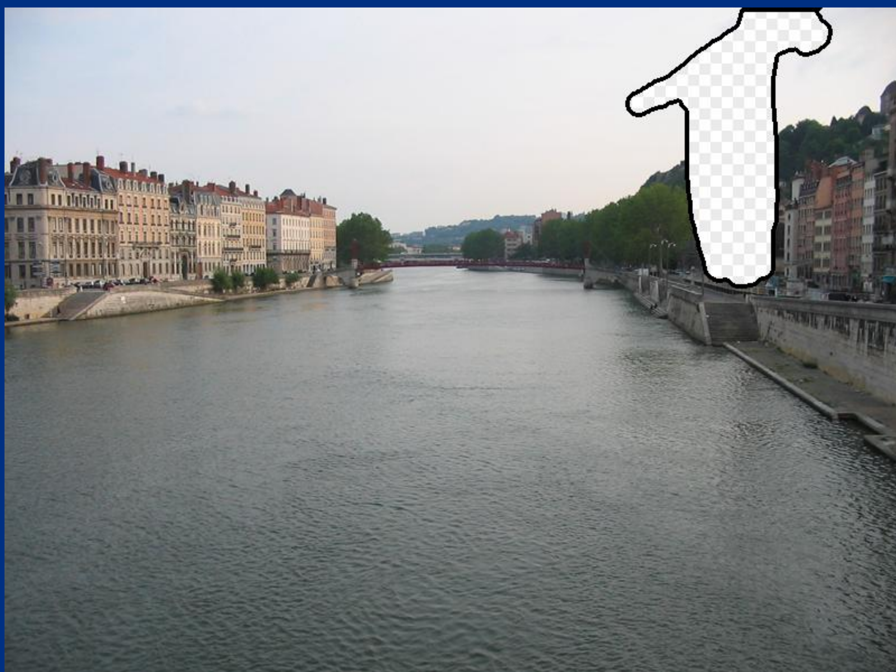
… 200 scene matches



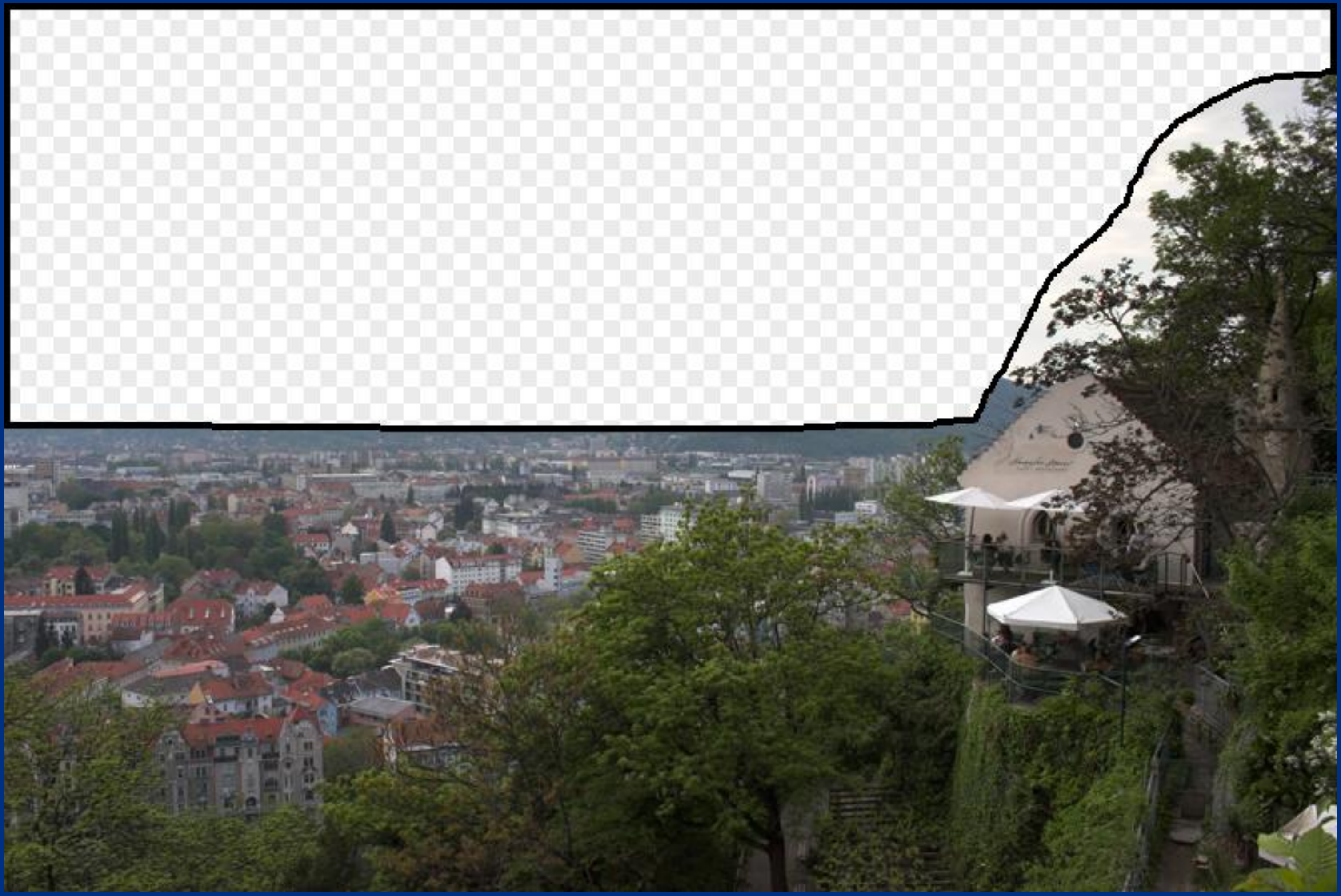
… 200 scene matches

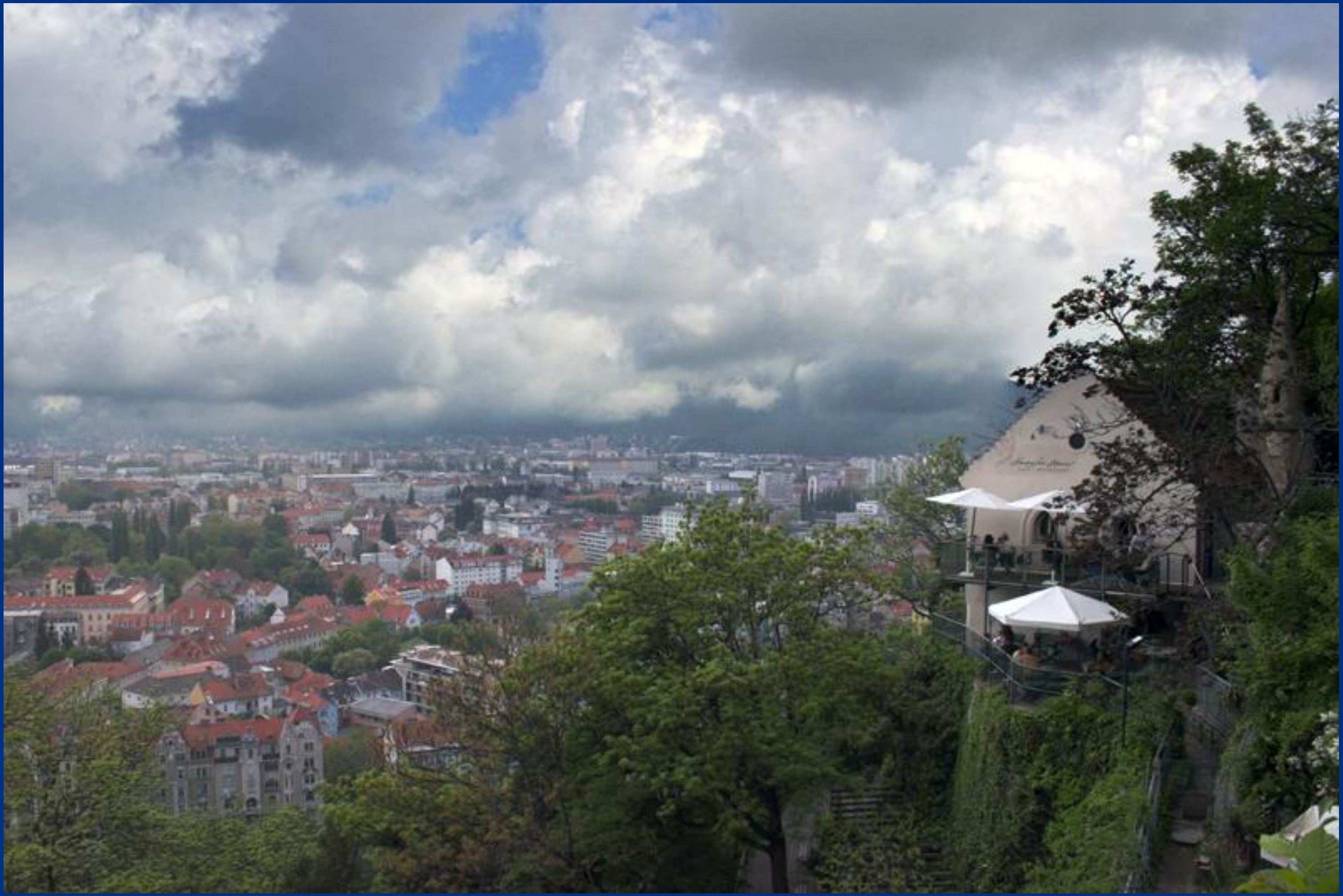


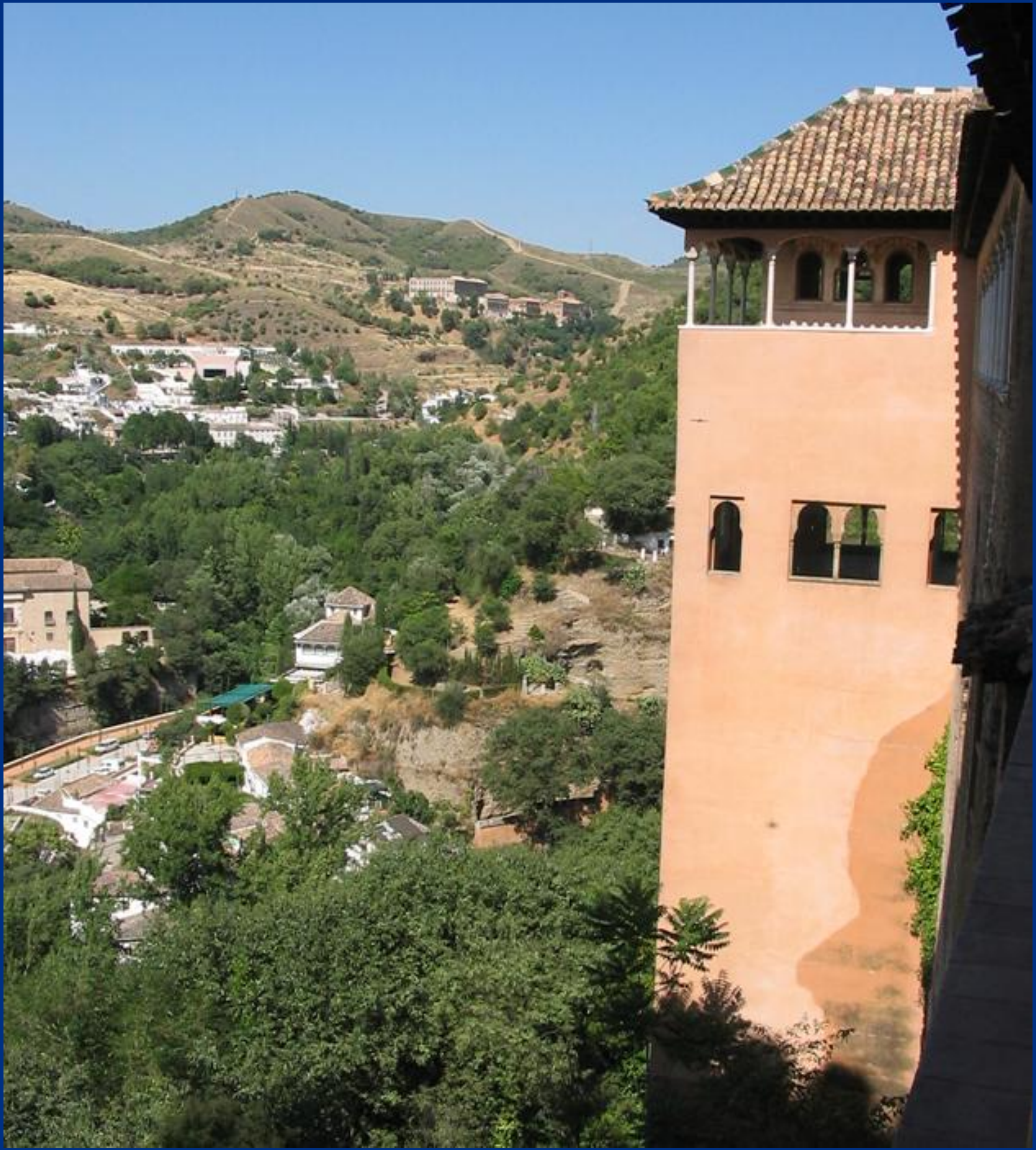
… 200 scene matches

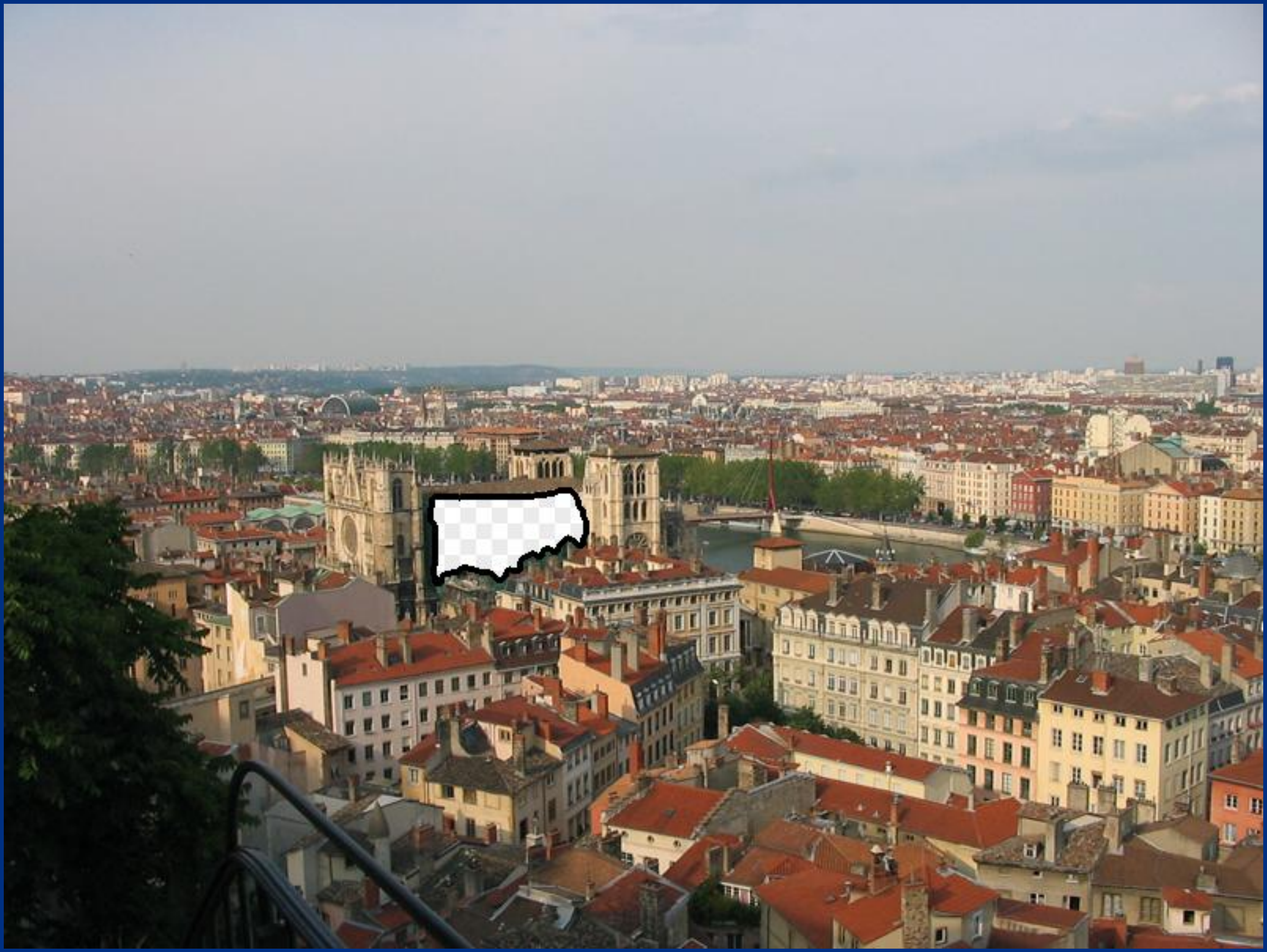


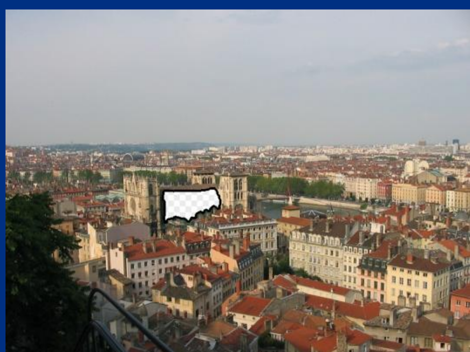

七五三詣り



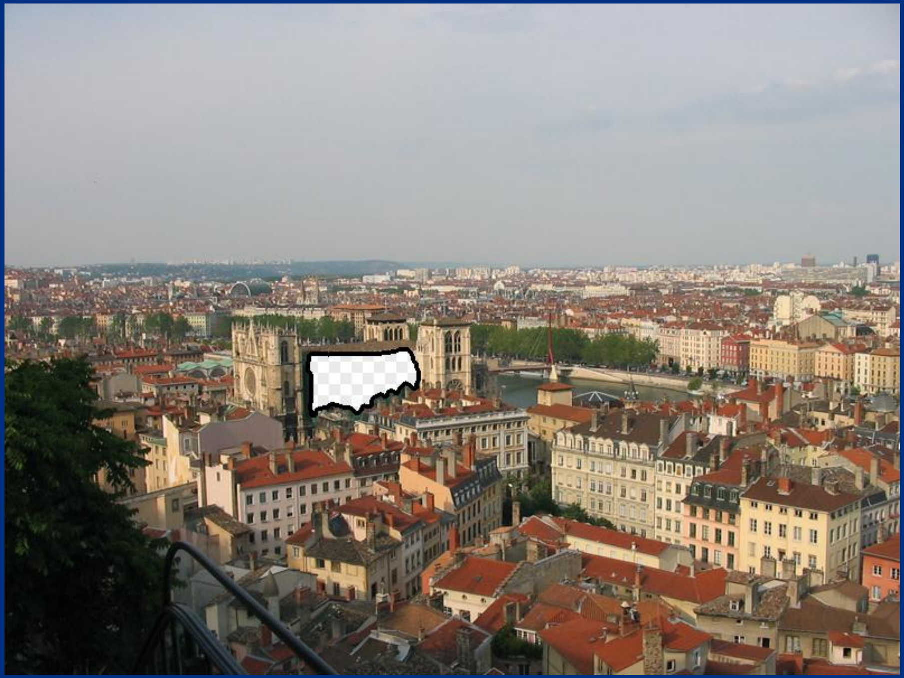








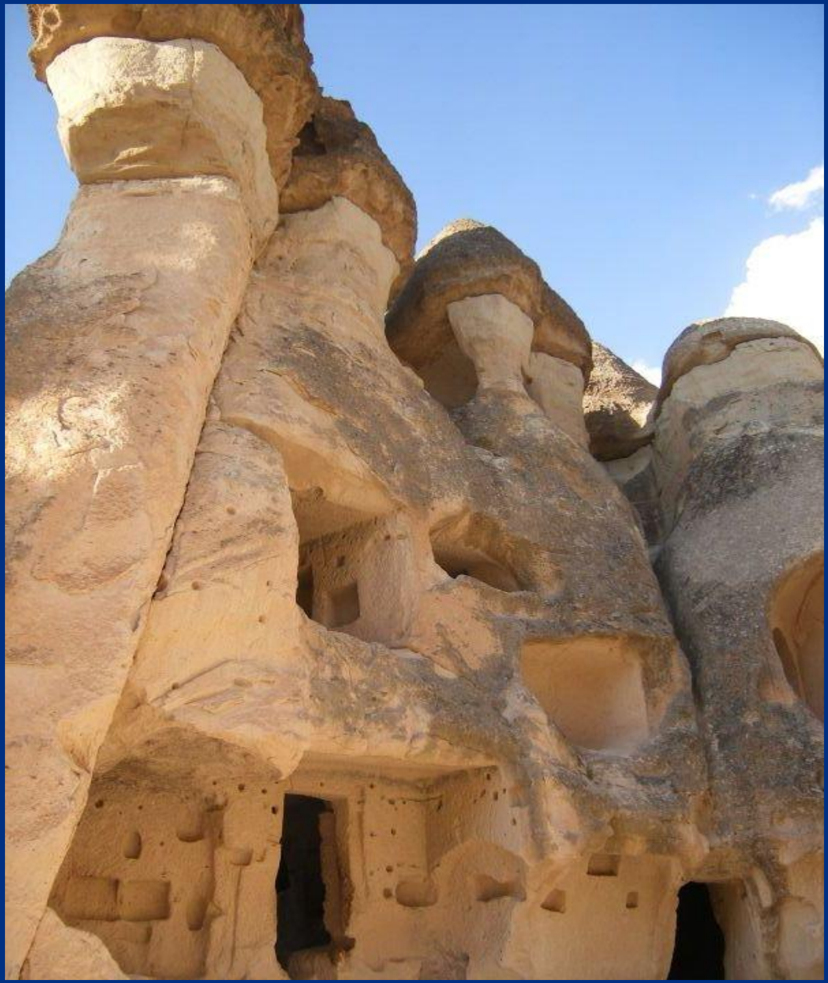
… 200 scene matches




Failures

Failures

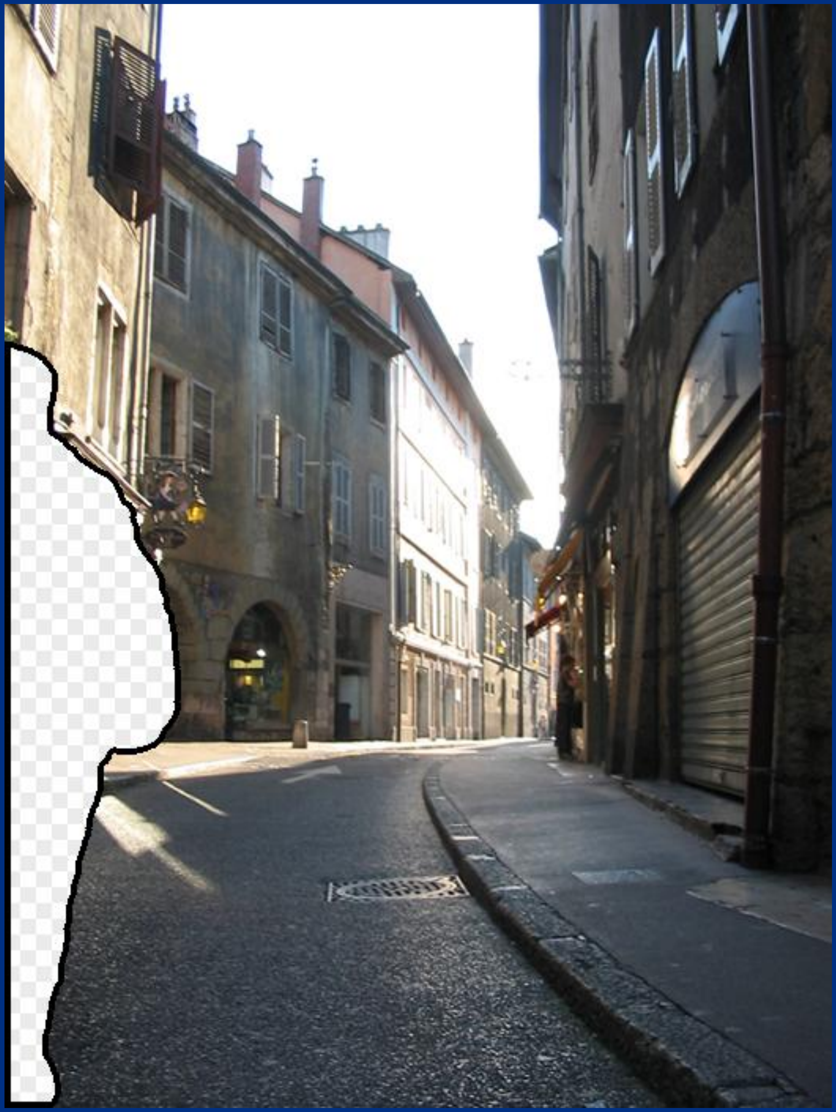
Failures



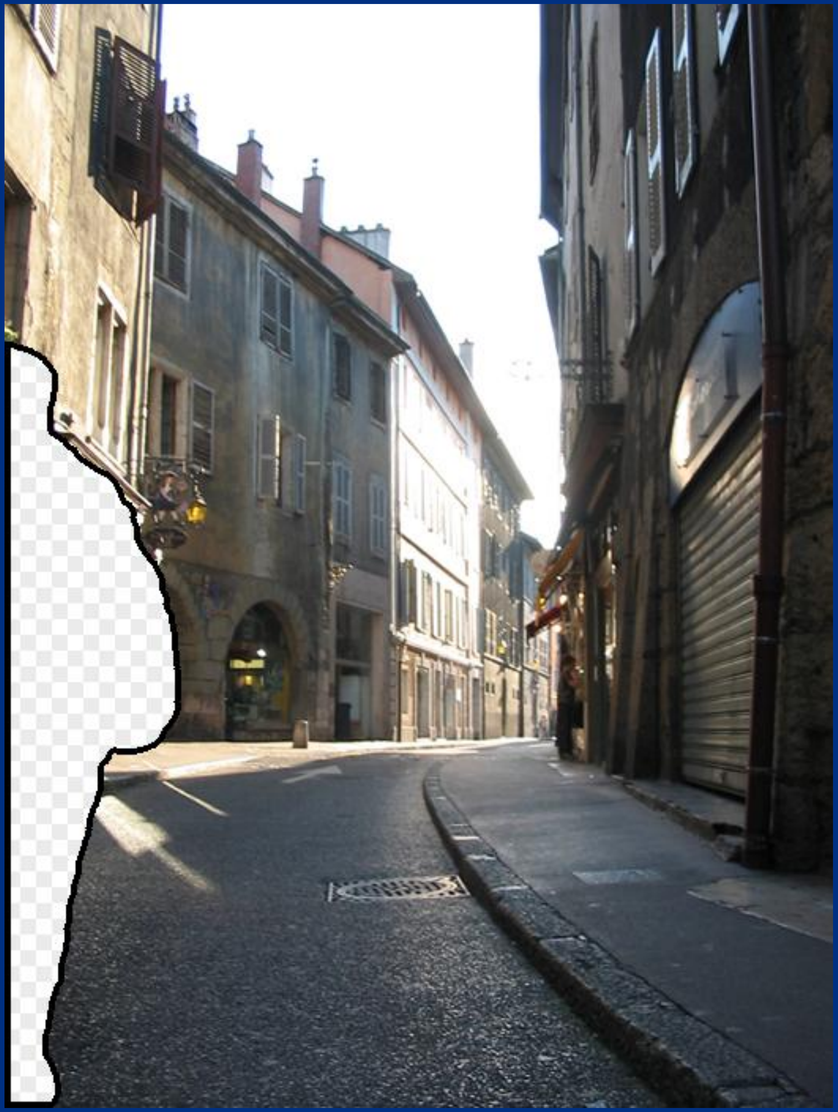
Failures

Failures

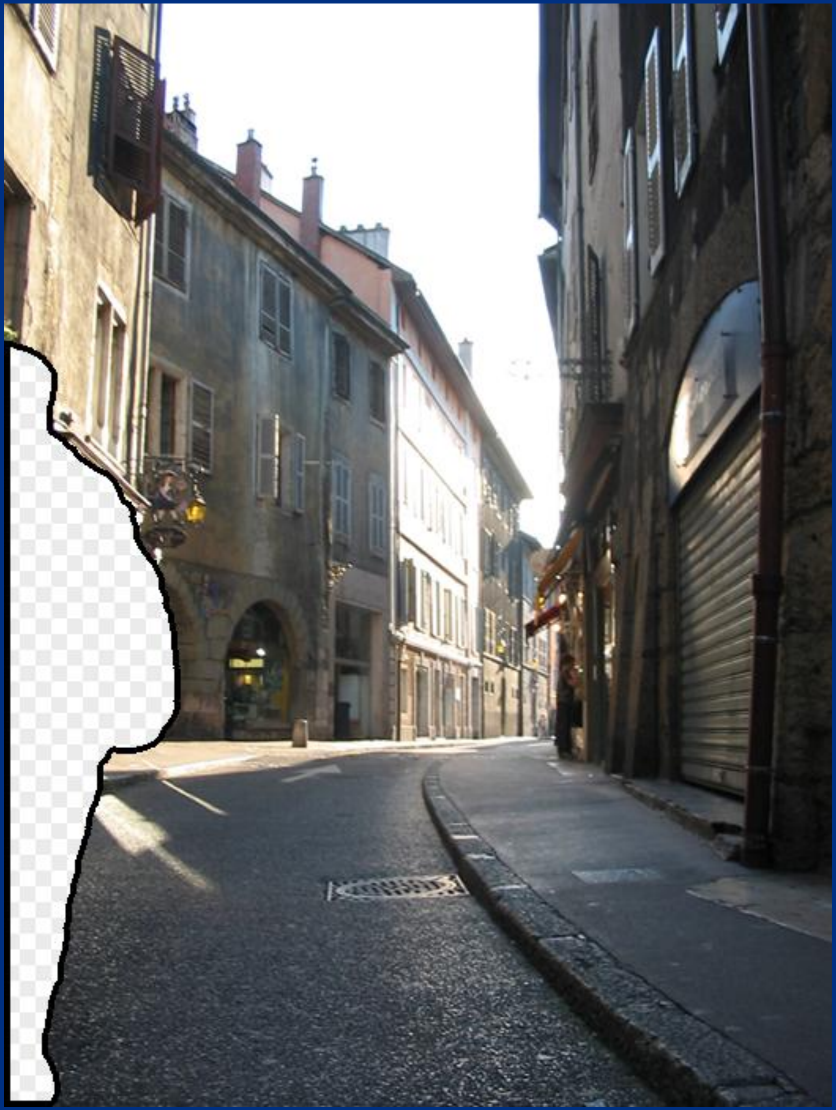
Failures



Failures

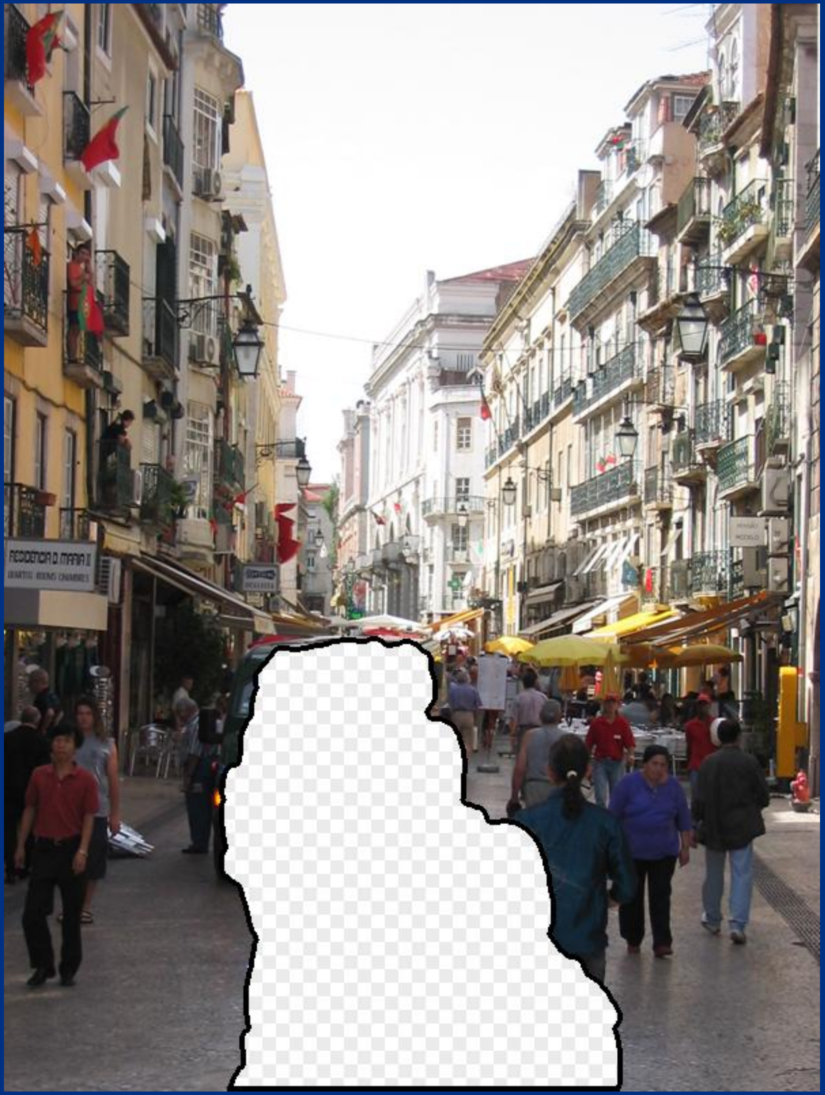


Failures



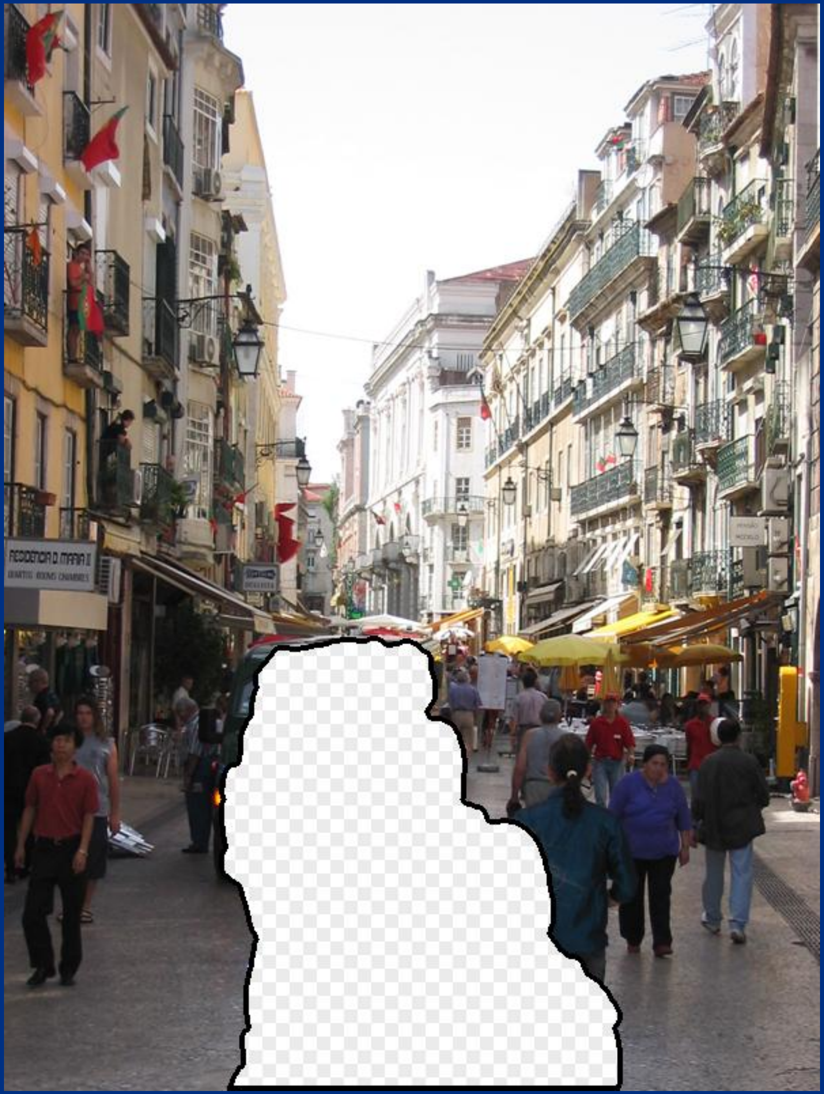
Failures

Failures



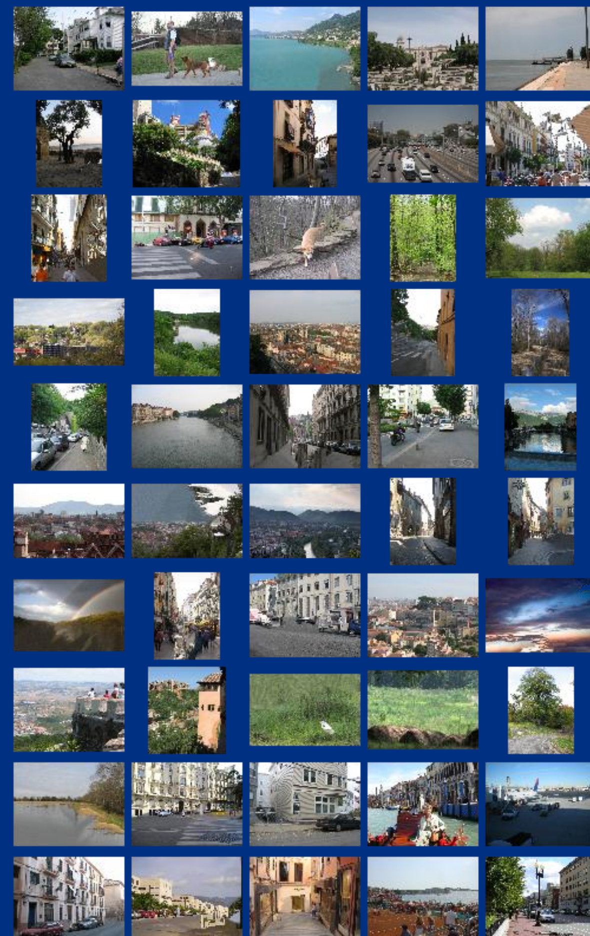
Failures

Failures



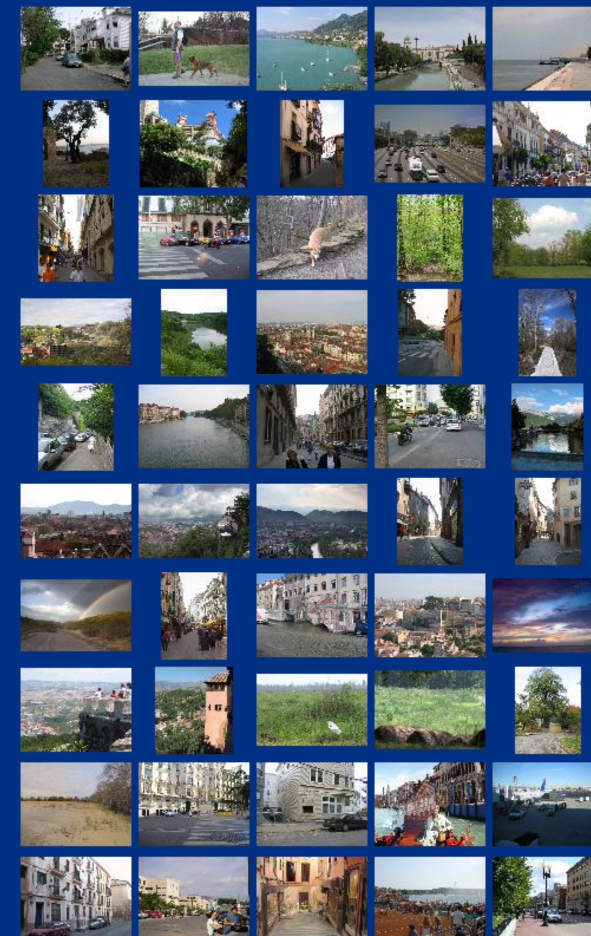
Evaluation

Original Images



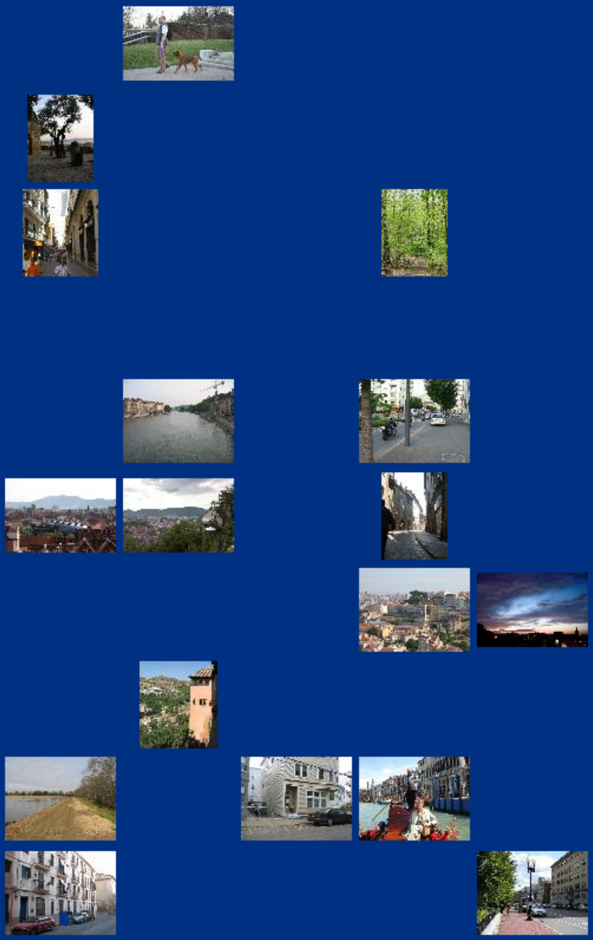
Criminisi et al.

Single result

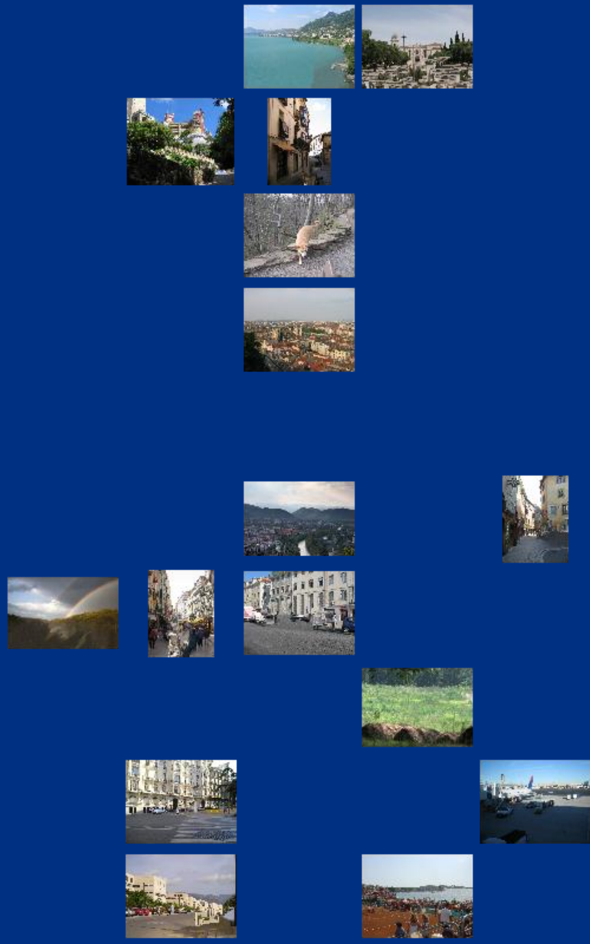


Scene Completion

Each result
selected from 20

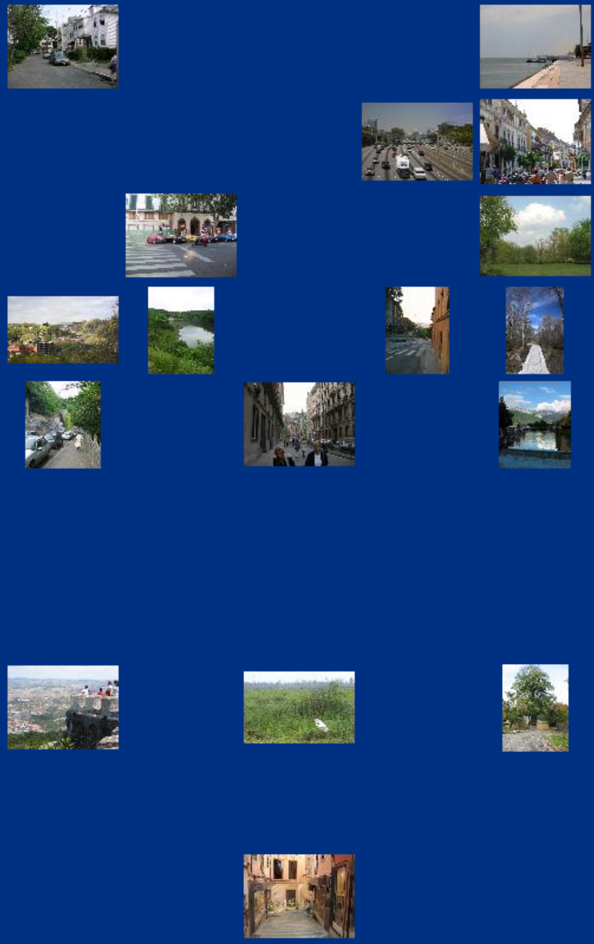


Original Images



Criminisi et al.

Single result



Scene Completion

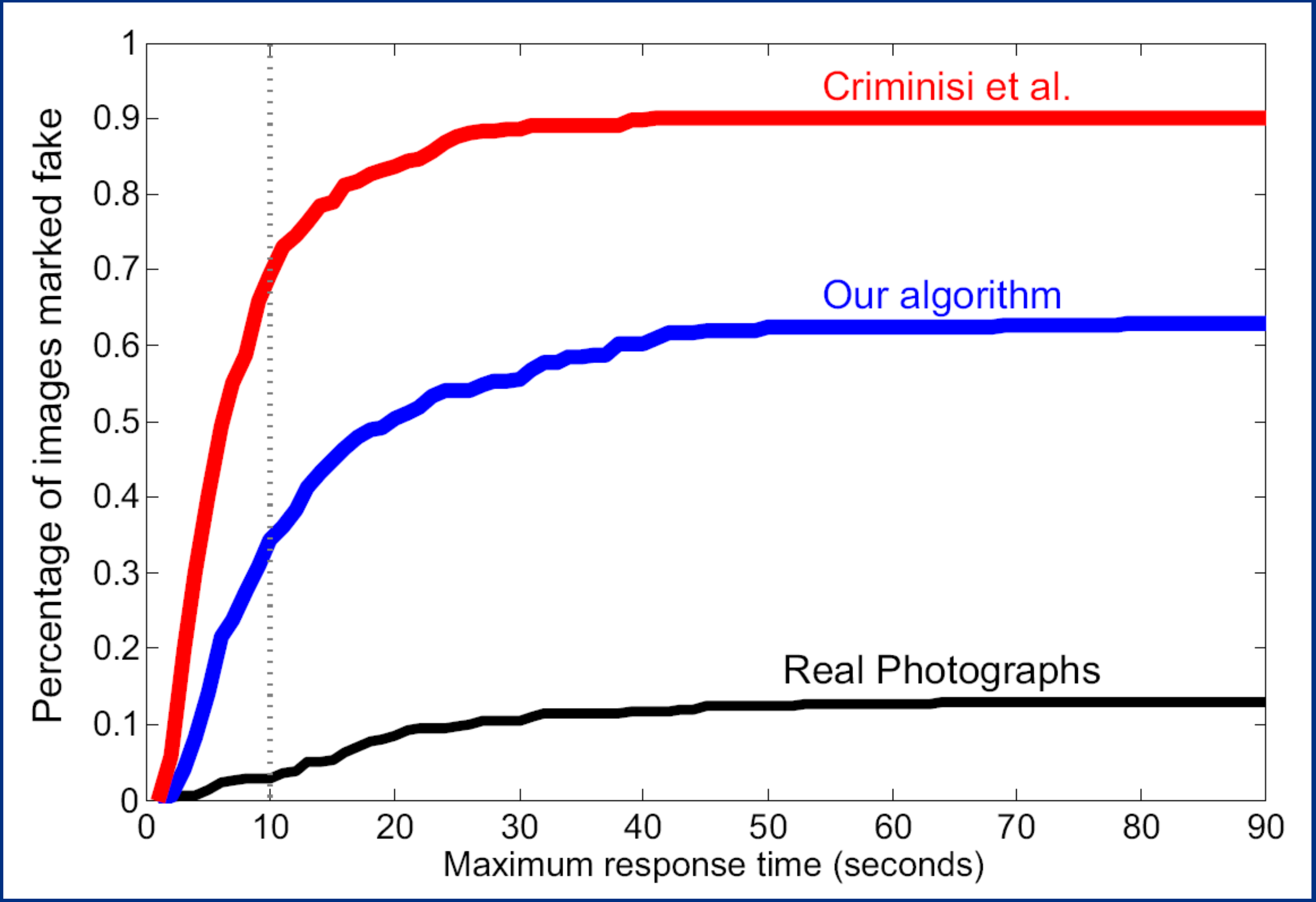
Each result
selected from 20

Real Image. This image has not been manipulated

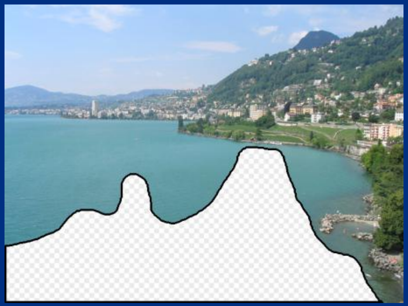
or

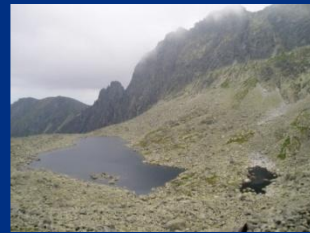
Fake Image. This image has been manipulated

User Study Results - 20 Participants

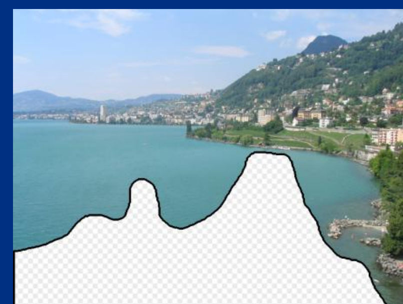


Why does it work?



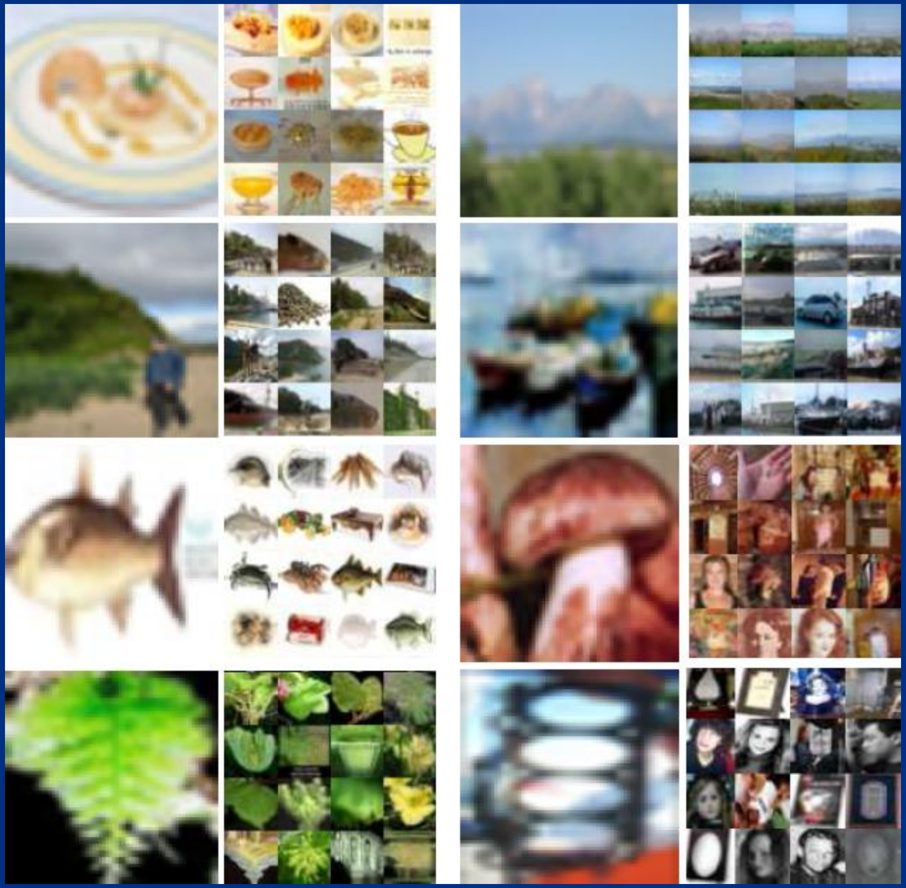







10 nearest neighbors from a collection of 20,000 images

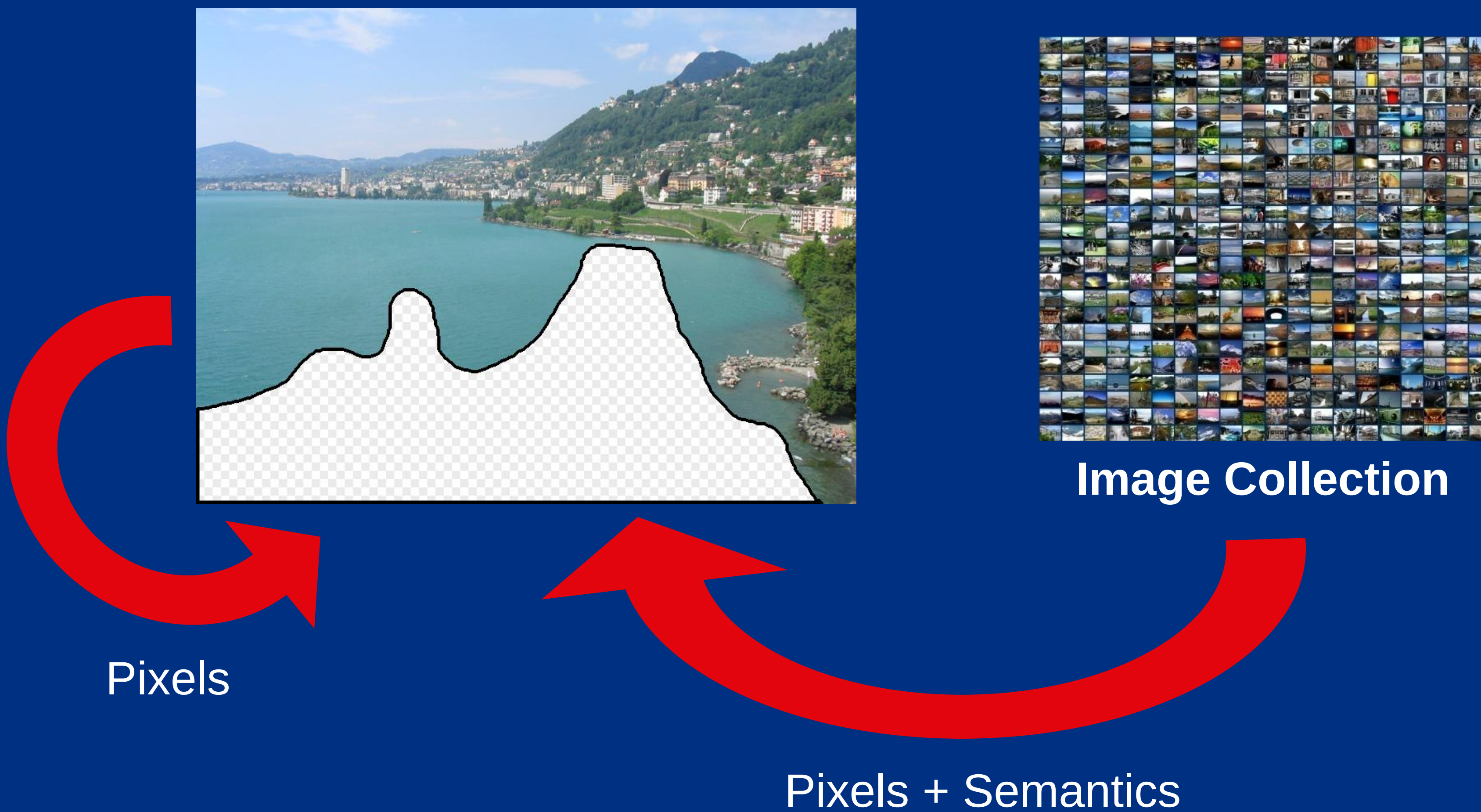
10 nearest neighbors from a
collection of 2 million images



Database of 70 Million 32x32 images

Torralba, Fergus, and Freeman. Tiny Images.
MIT-CSAIL-TR-2007-024. 2007.

The Small Picture



Hybrid Solution?

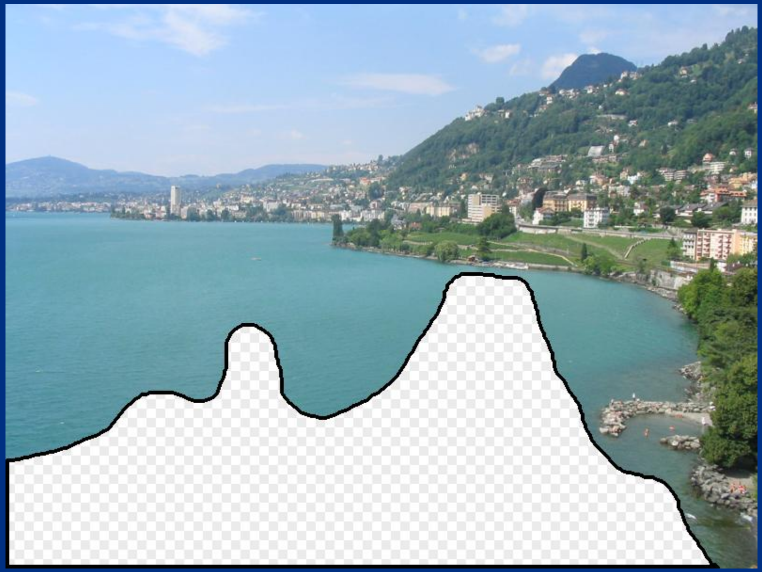


Image Collection

Pixels

Semantics

The Big Picture

Sky, Water, Hills, Beach,
Sunny, mid-day

<u>Brute-force</u> Image Understanding

80 Million Tiny Images

Massachusetts Institute of Technology

Antonio Torralba

Rob Fergus

William T. Freeman

New York University

Admin

- HW4 due on Thursday 12th May
- This is a hard deadline!
- The TA has to grade the assignment by Saturday so I can turn in grades

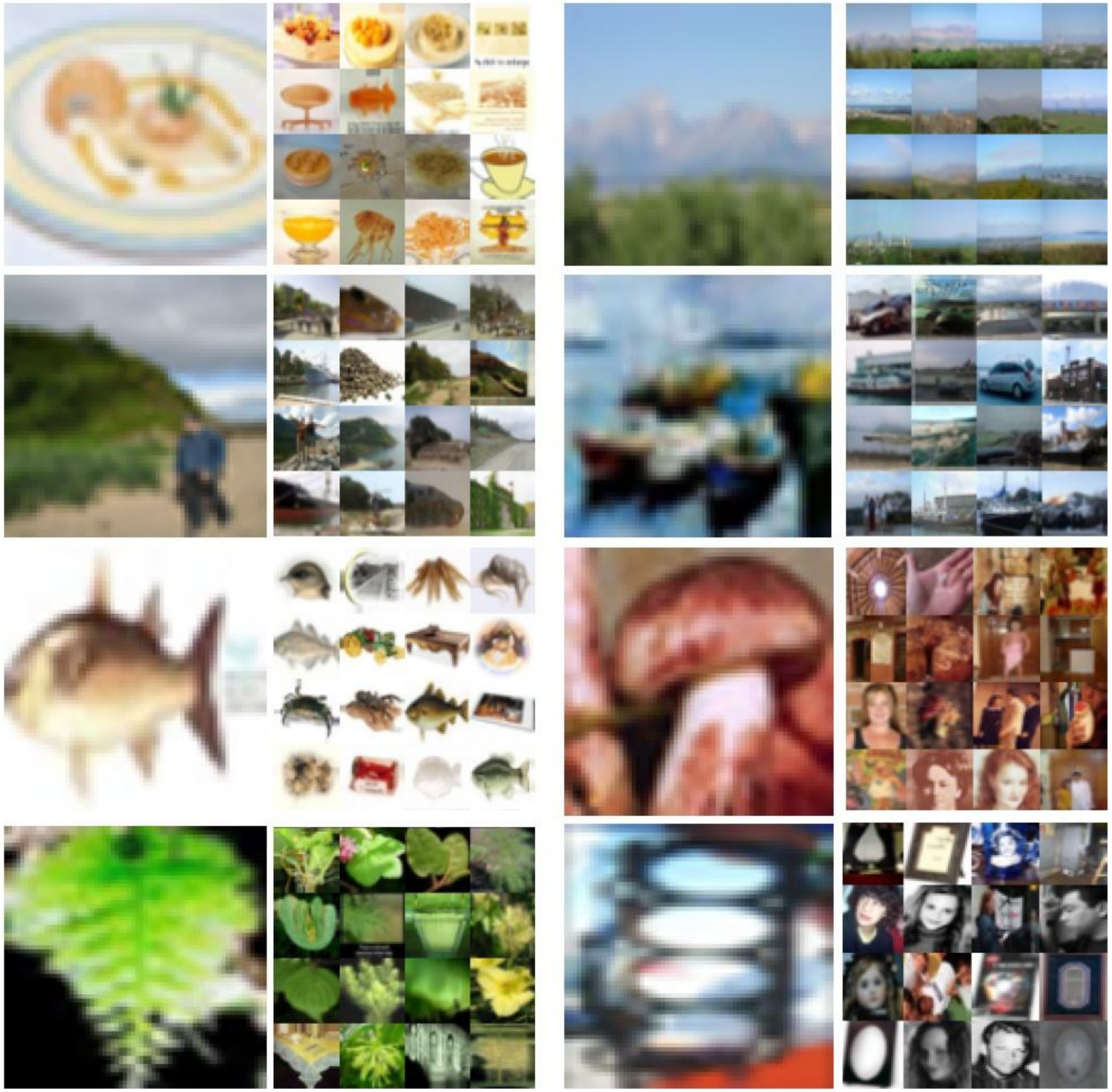
Overview

- Non-parametric approach to category-level recognition
- Dataset of 80 million images from Internet

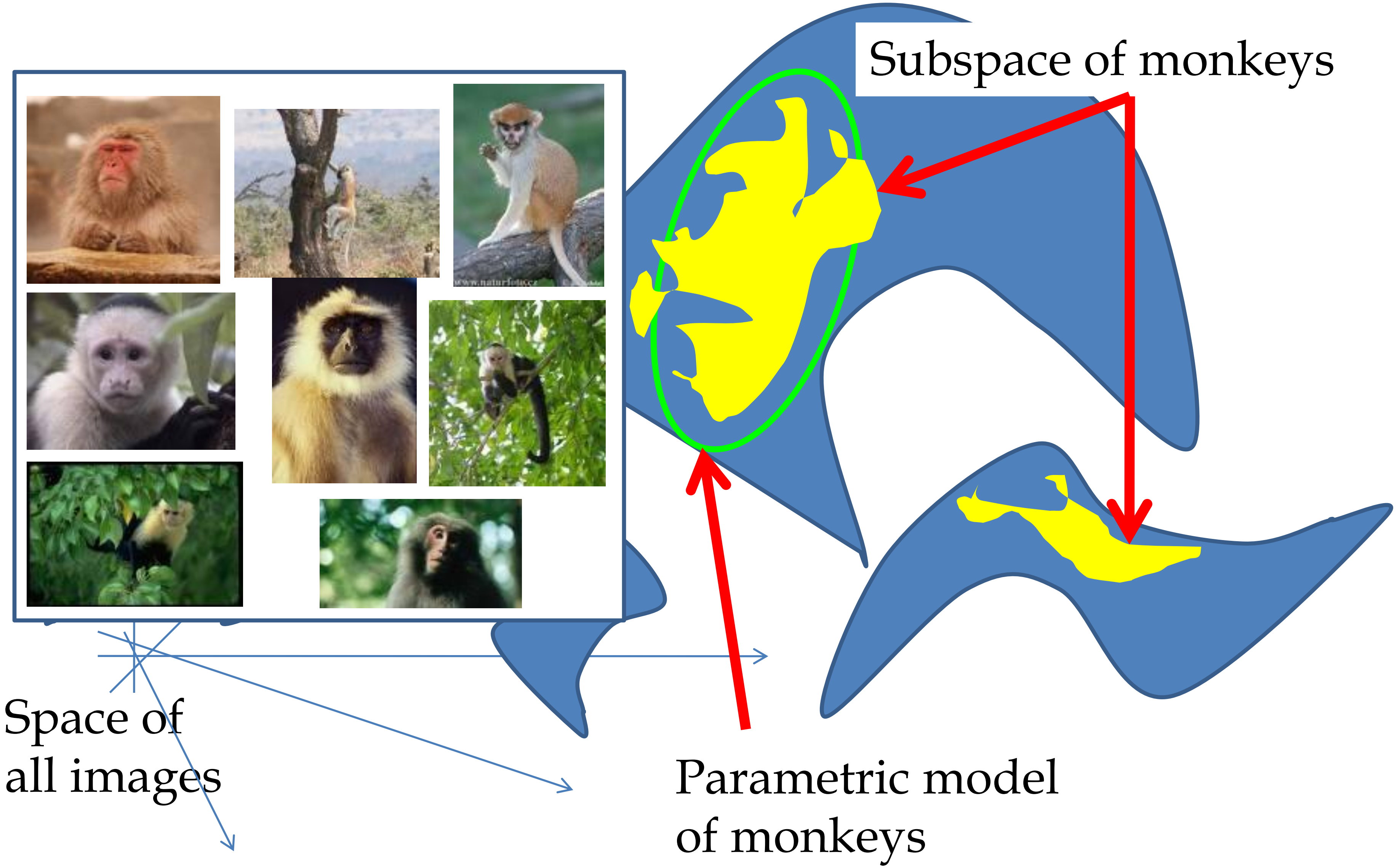
- Use very low resolution images (32x32 color)

Overview

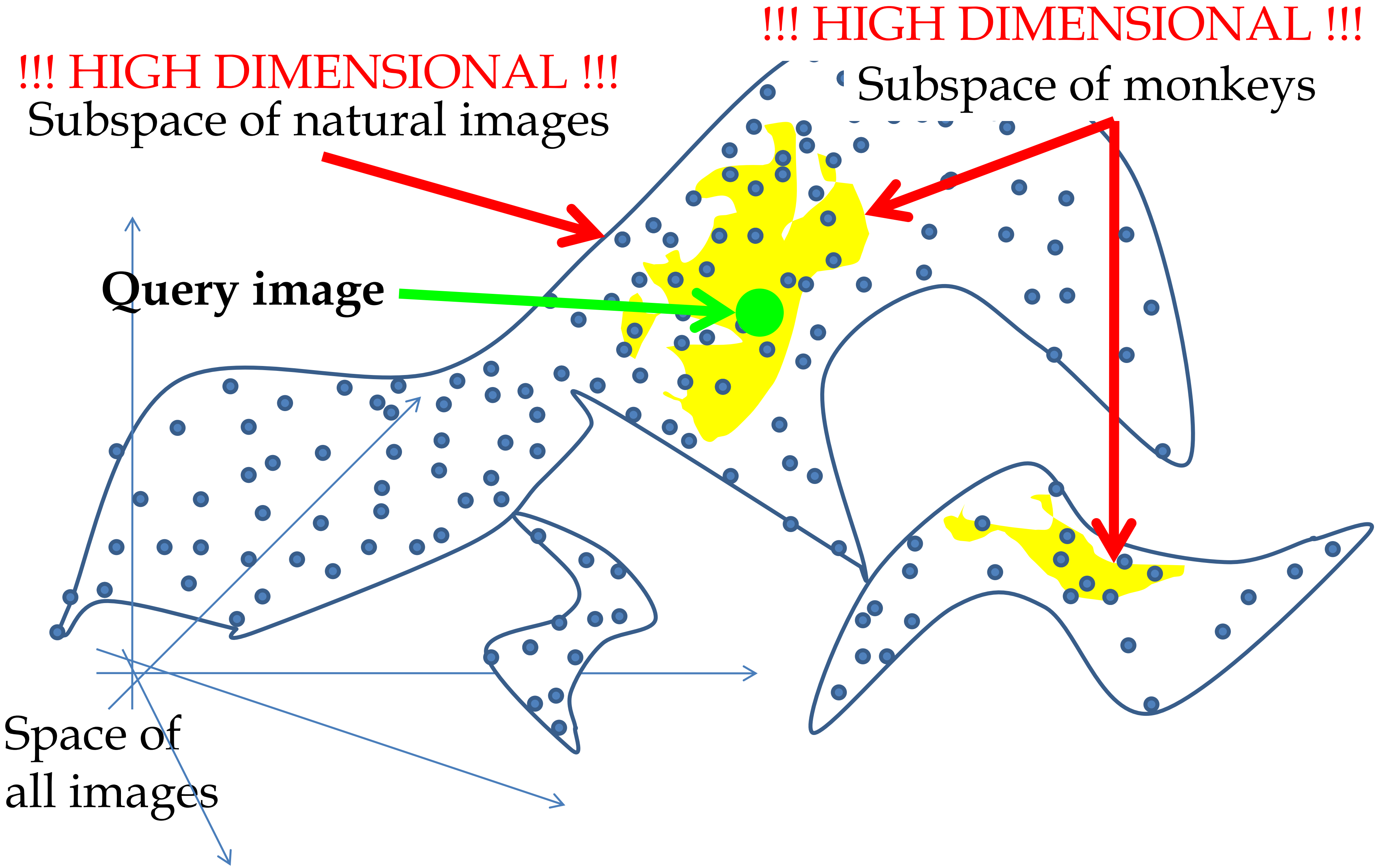
- Use simple algorithms: nearest neighbors



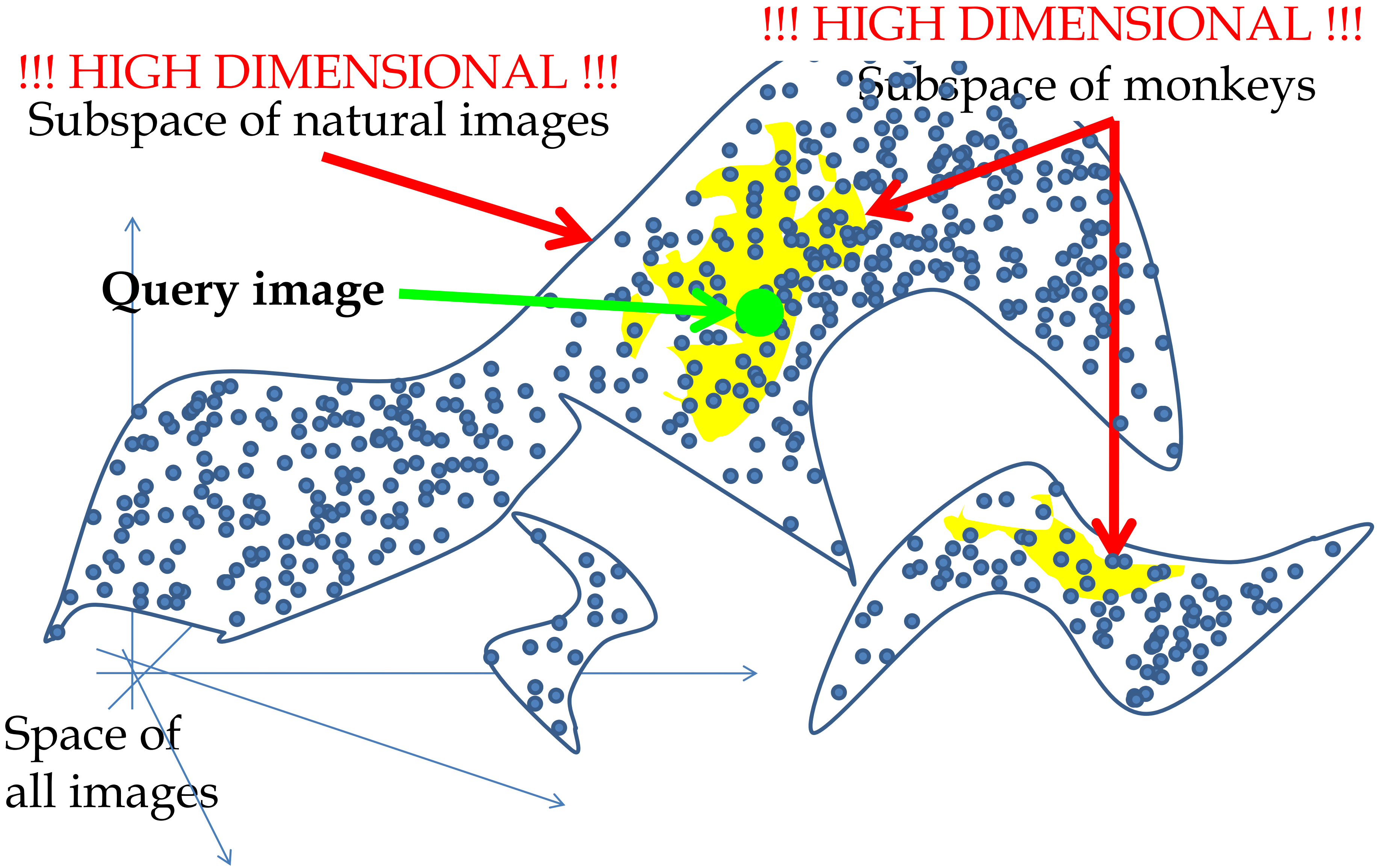
Motivation


Subspace of monkeys
Space of
all images
Parametric model
of monkeys

Non-parametric Approach

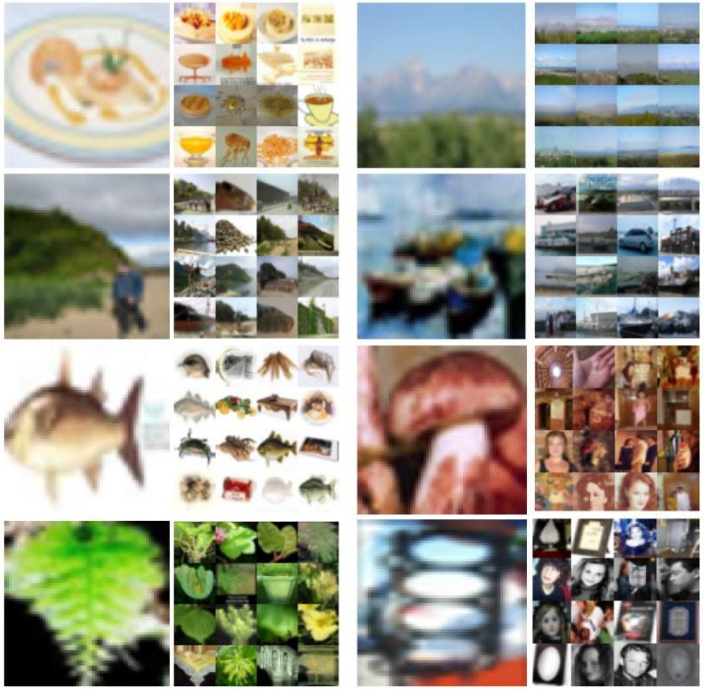


Non-parametric Approach



Non-parametric Classifier

- Nearest-neighbors
- For each query, obtain sibling set (neighbors)
- 3 different types of distance metric
- Hand-designed, use whole image



Metric 1 - D_{ssd}

- Sum of squared differences (SSD)

$$D^2_{ssd} = \sum_{x,y,c} \left[\text{Image 1} - \text{Image 2} \right]^2$$

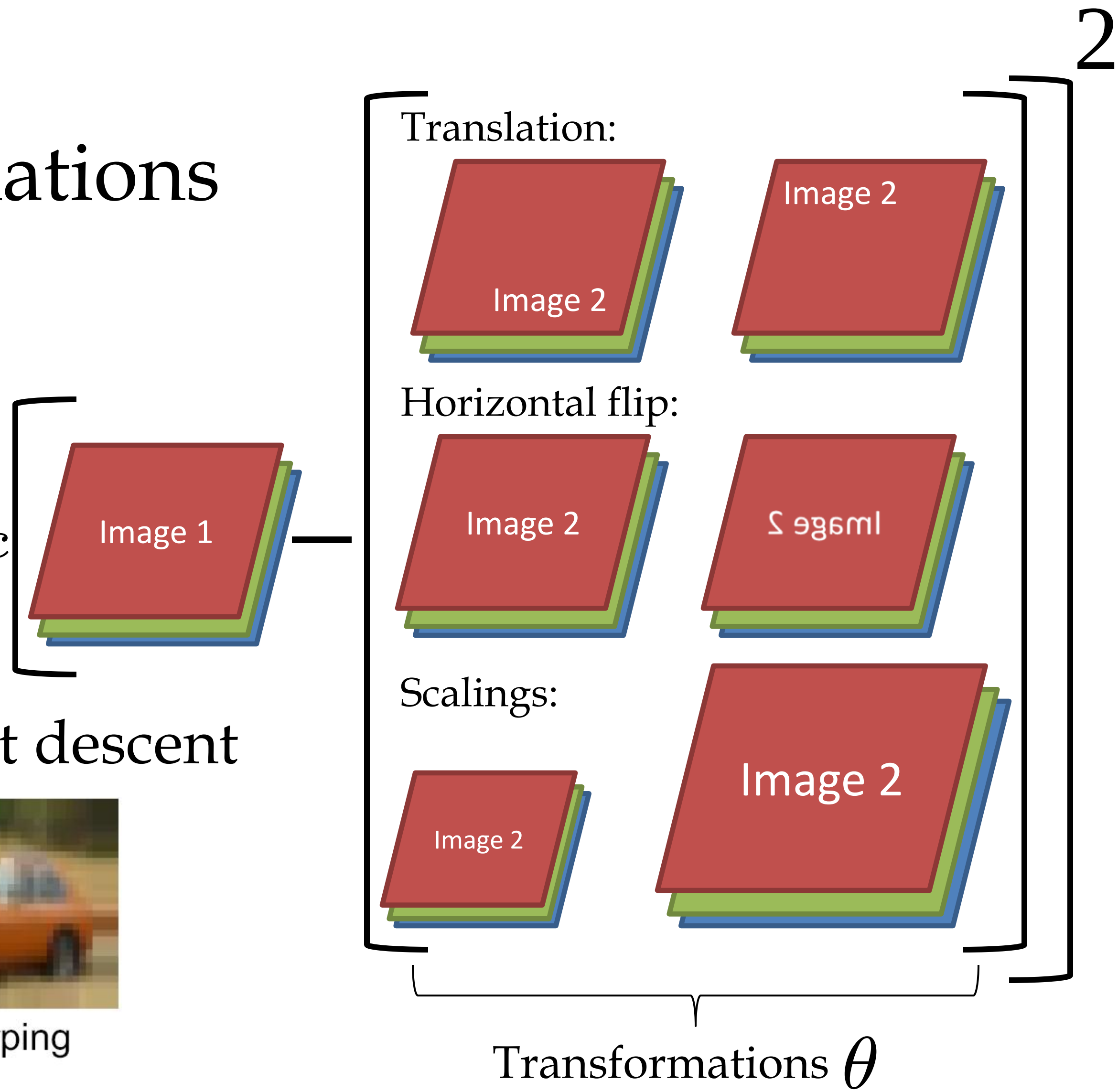
To give invariance to illumination:
Each image normalized to
be zero mean, unit variance

Target Neighbor

Metric 2 - D_{warp}

- SSD but allow small transformations

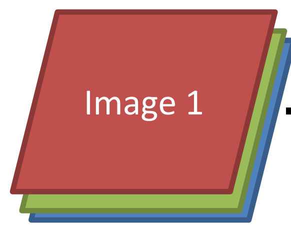
$$D^2_{warp} = \min_{\theta} \sum_{x,y,c}$$

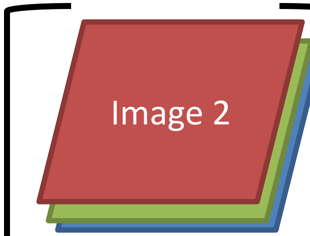


Find min using gradient descent

Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

$D^2_{shift} = \sum_{x,y,c} \Big[$ 

$-$ Transformed θ $\Big[$ 

$\Big]\Big]^2$

Start with warped version of image 2, as per D_{warp}

Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

$$D^2_{shift} = \sum_{x,y,c} \Bigg[$$

$$- \underset{\theta}{\text{Transformed}} \Big[$$

$$\Big] \Bigg]^2$$

Start with warped version of image 2, as per D_{warp}

Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

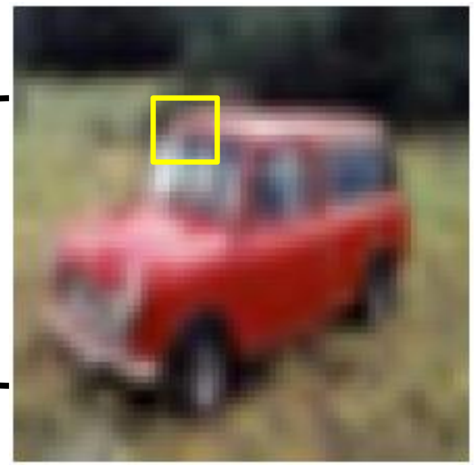
$D^2_{shift} = \sum_{x,y,c} \Big[$  $-$  $\Big]^2$

Start with warped version of image 2, as per D_{warp}

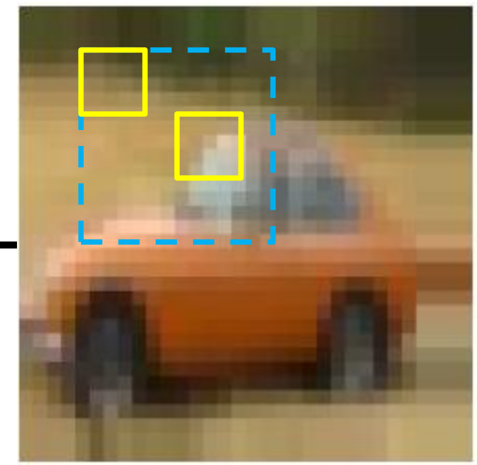
Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

$$D^2_{shift} = \min_{\substack{\text{Local}\\\text{sub-window}}} \sum_{x,y,c} \Big[$$



$$-$$

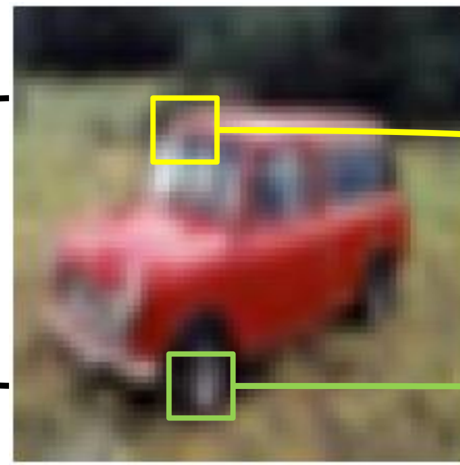


$$\Big]^2$$

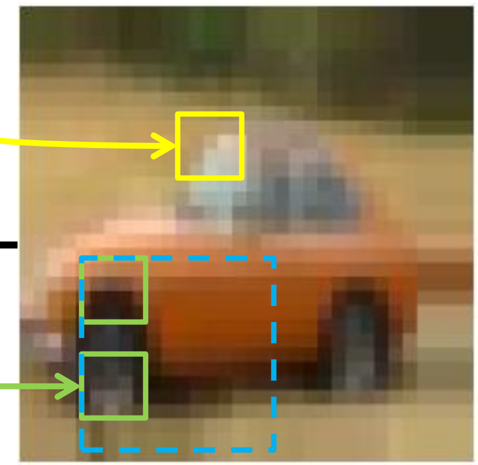
Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

$$D^2_{shift} = \min_{\substack{\text{Local} \\ \text{sub-window}}} \sum_{x,y,c} \Big[$$



$-$



$$\Big]^2$$

- Quick since images are so small

Metric 3 - D_{shift}

- As per Warping but also allow sub-window shifts

$$D^2_{shift} = \min_{\substack{\text{Local}\\ \text{sub-window}}} \sum_{x,y,c} \Bigg[$$

$-$

$$\Bigg]^2$$

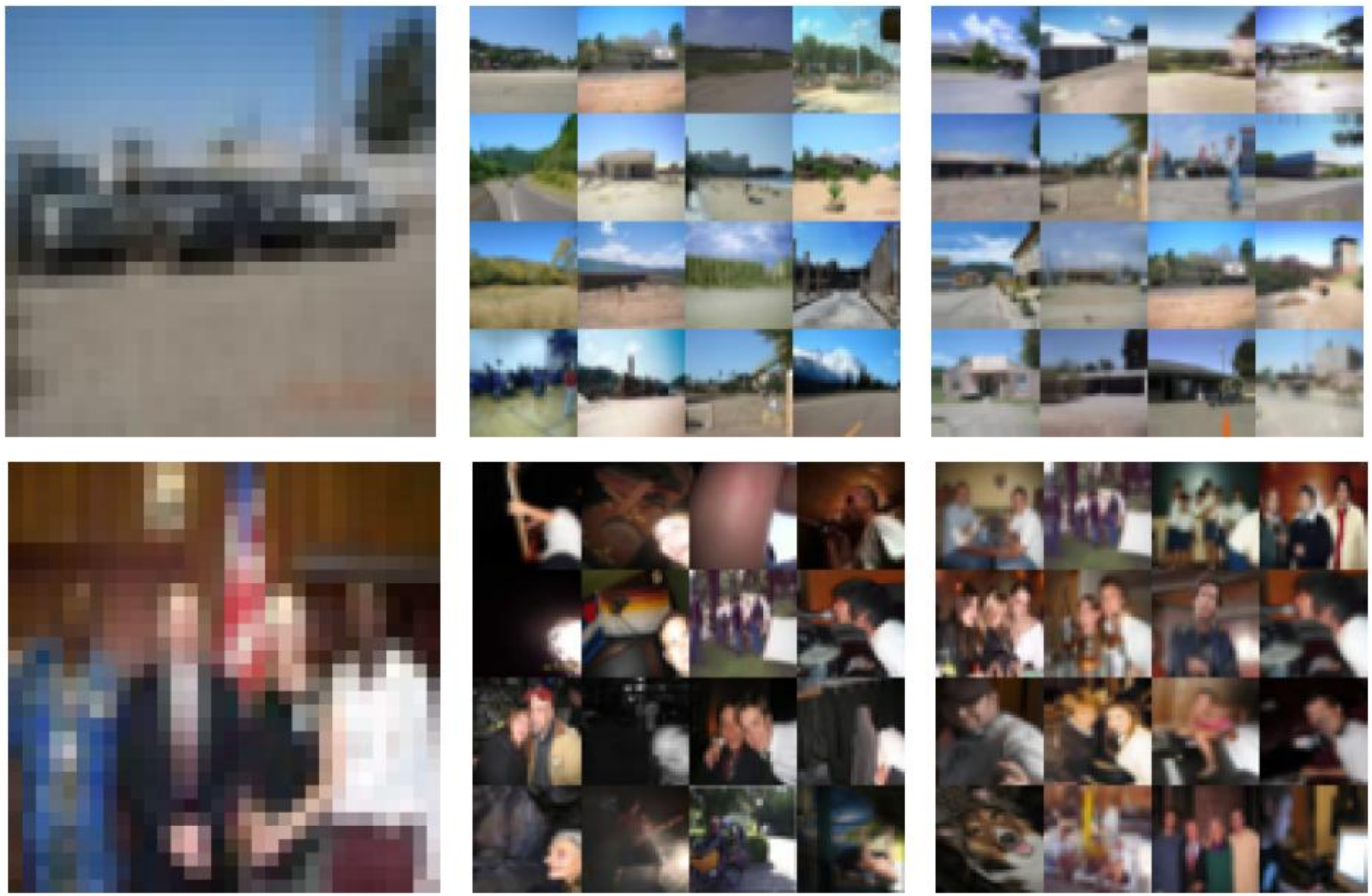
Tried various sizes of sub-window

→ 1x1 (i.e. single pixel) worked best

Comparison of metrics

Sibling Sets with Different Metrics

- Sibling set is 50 images

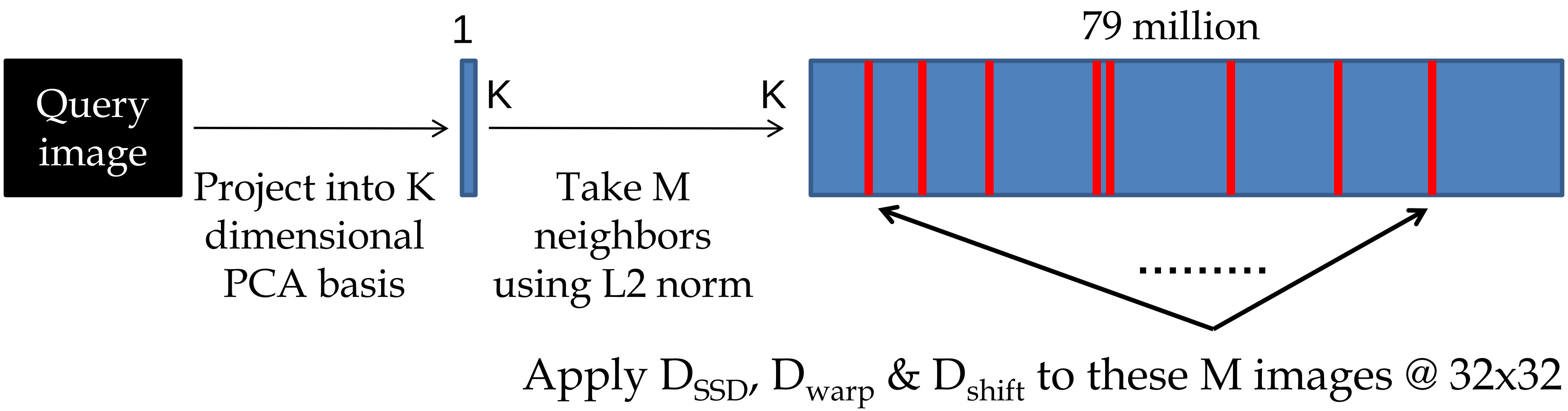


D_{ssd} D_{shift}

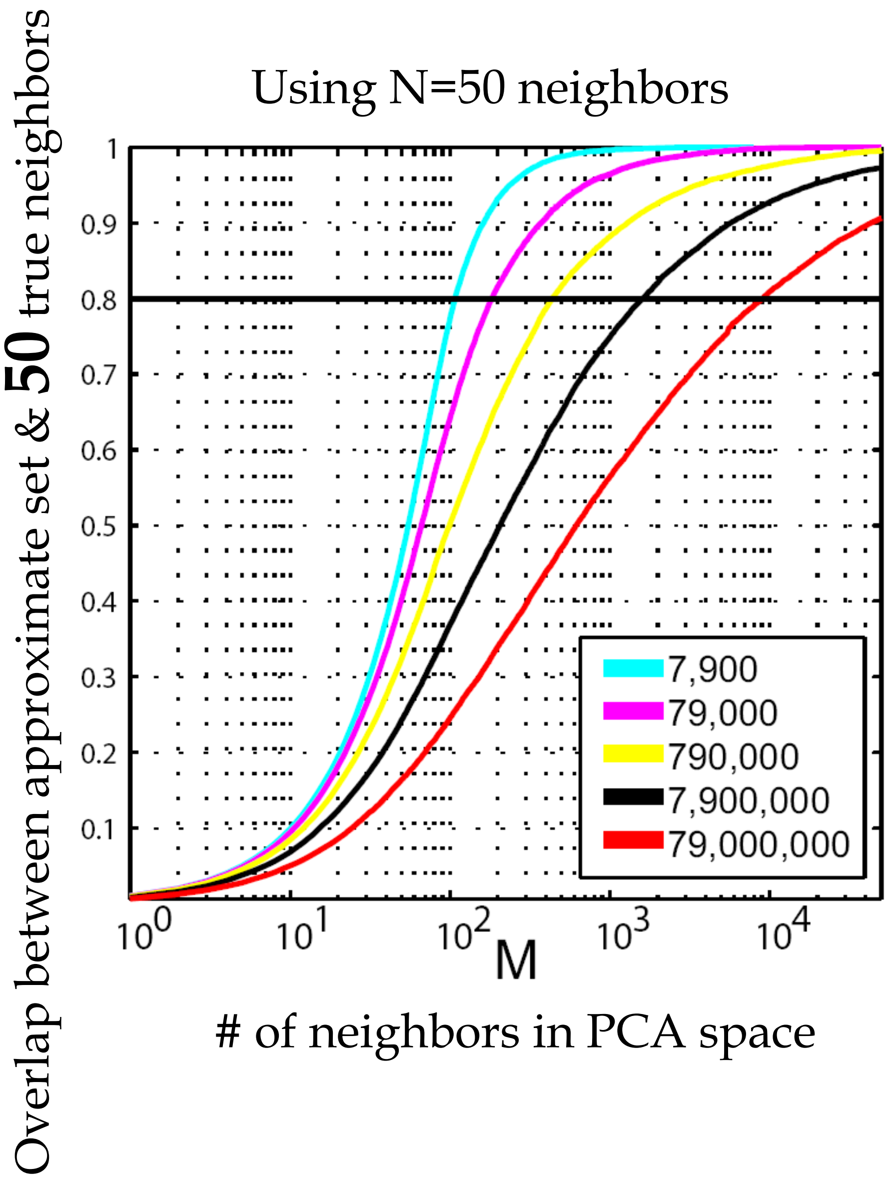
Approximate D_{ssd}

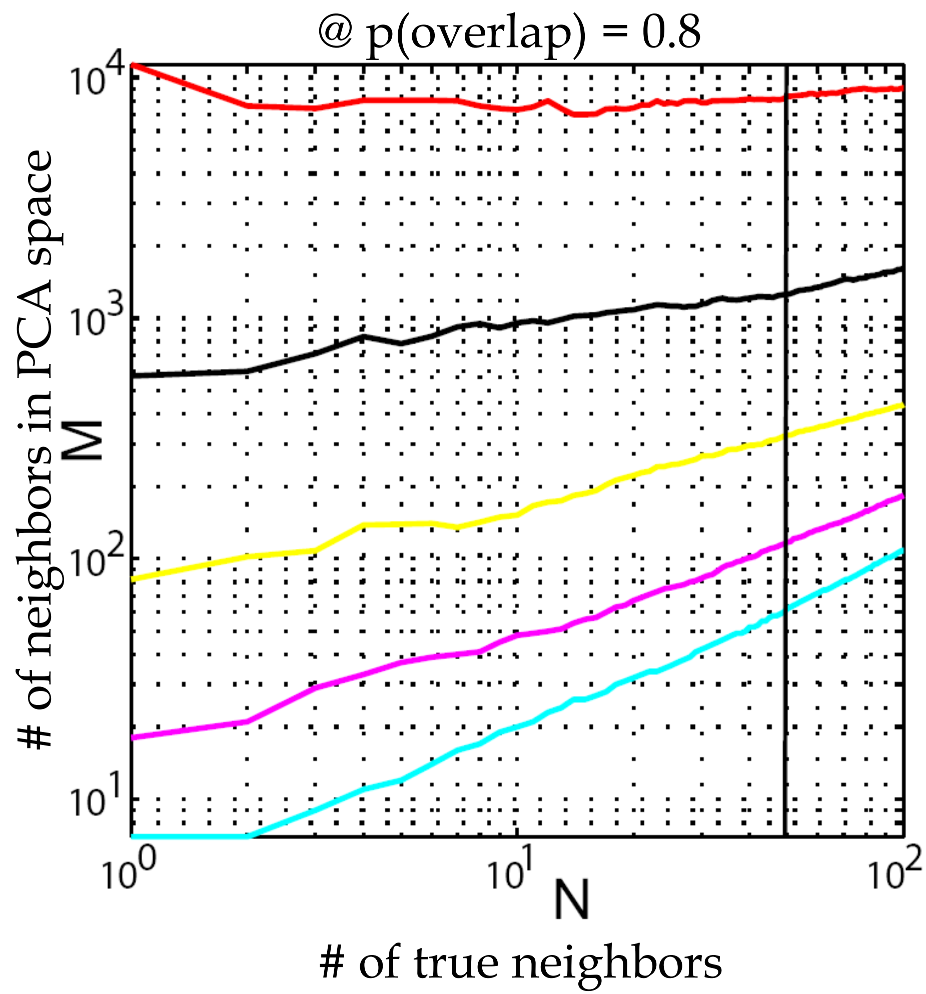
- Exact distance metrics are too expensive to apply to all 79 million images

- Use approximate scheme based on taking first K=19 principal components

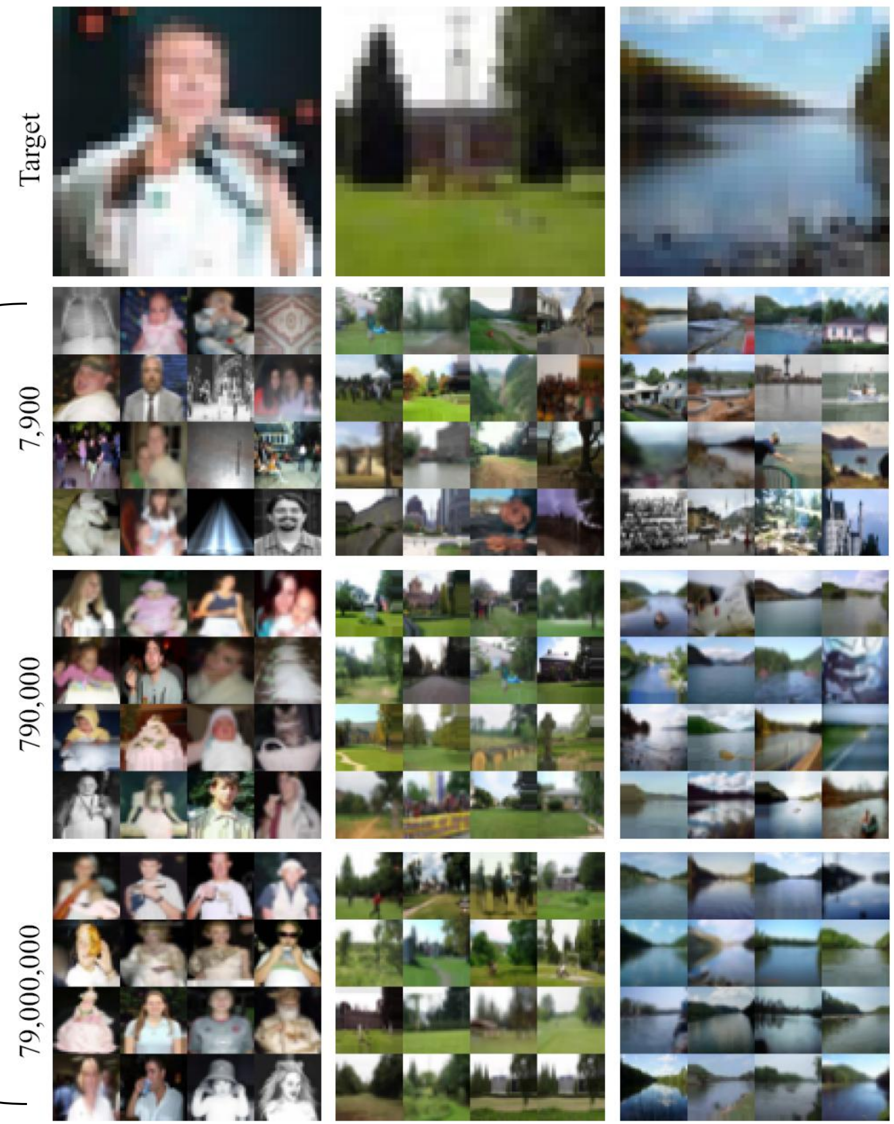


Exact SSD vs Approximate SSD





Quality of Sibling Set using D_{shift}



Size of dataset

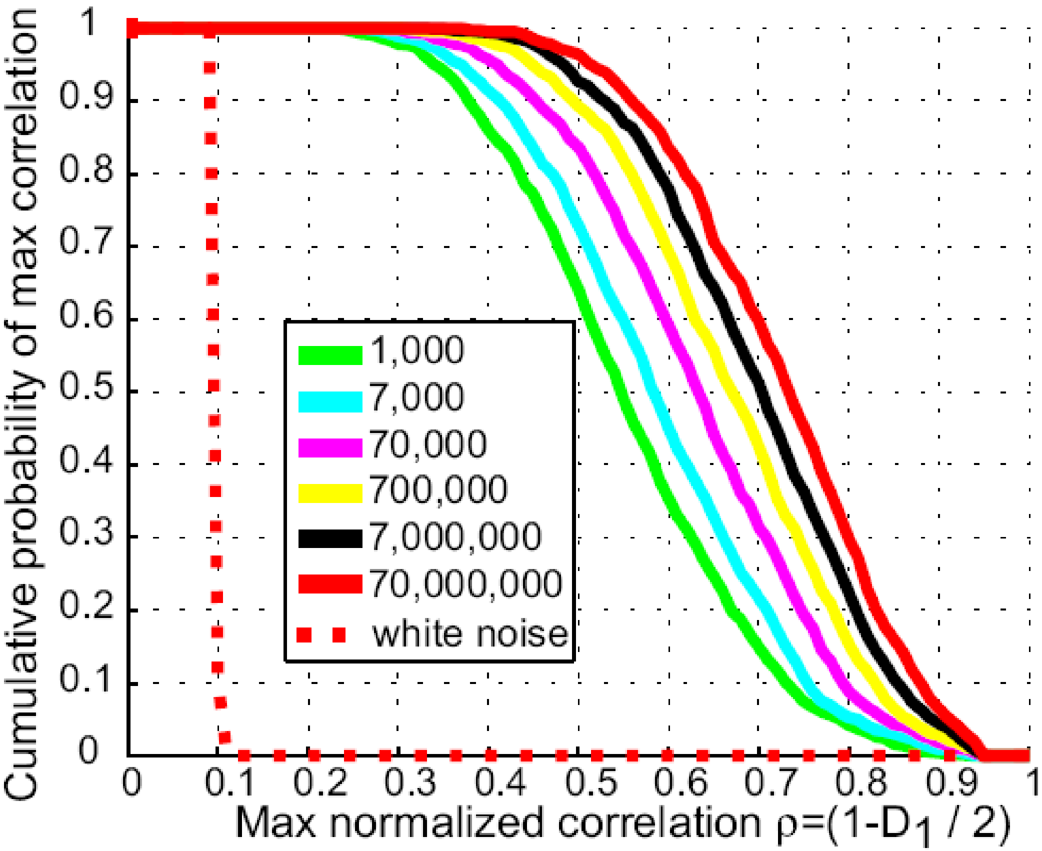
10^5

10^6

10^8

Exploring the Sub-Space of Natural Images

How Many Images Are There?

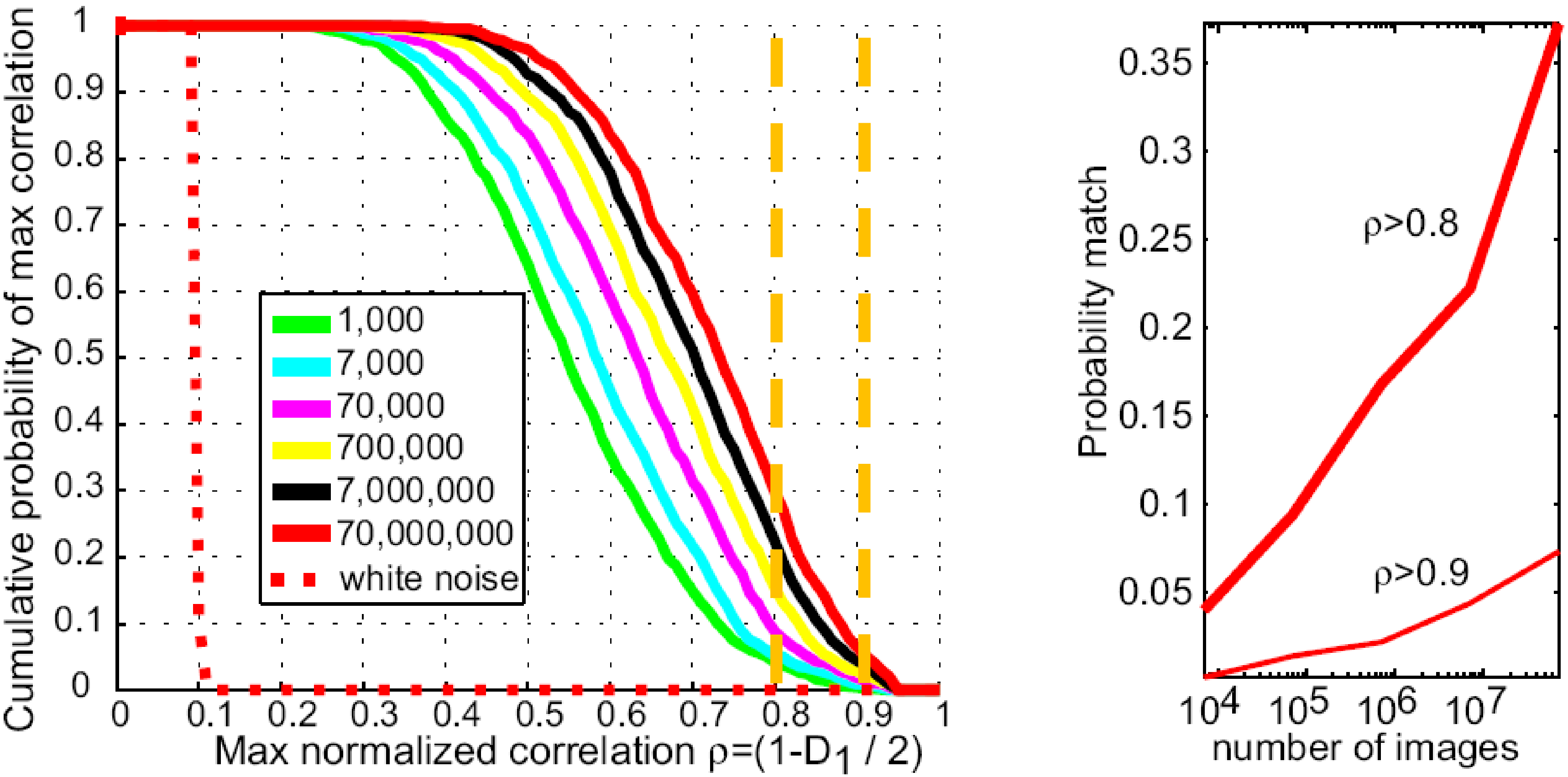


Note: $D_1 = D_{SSD}$

Examples

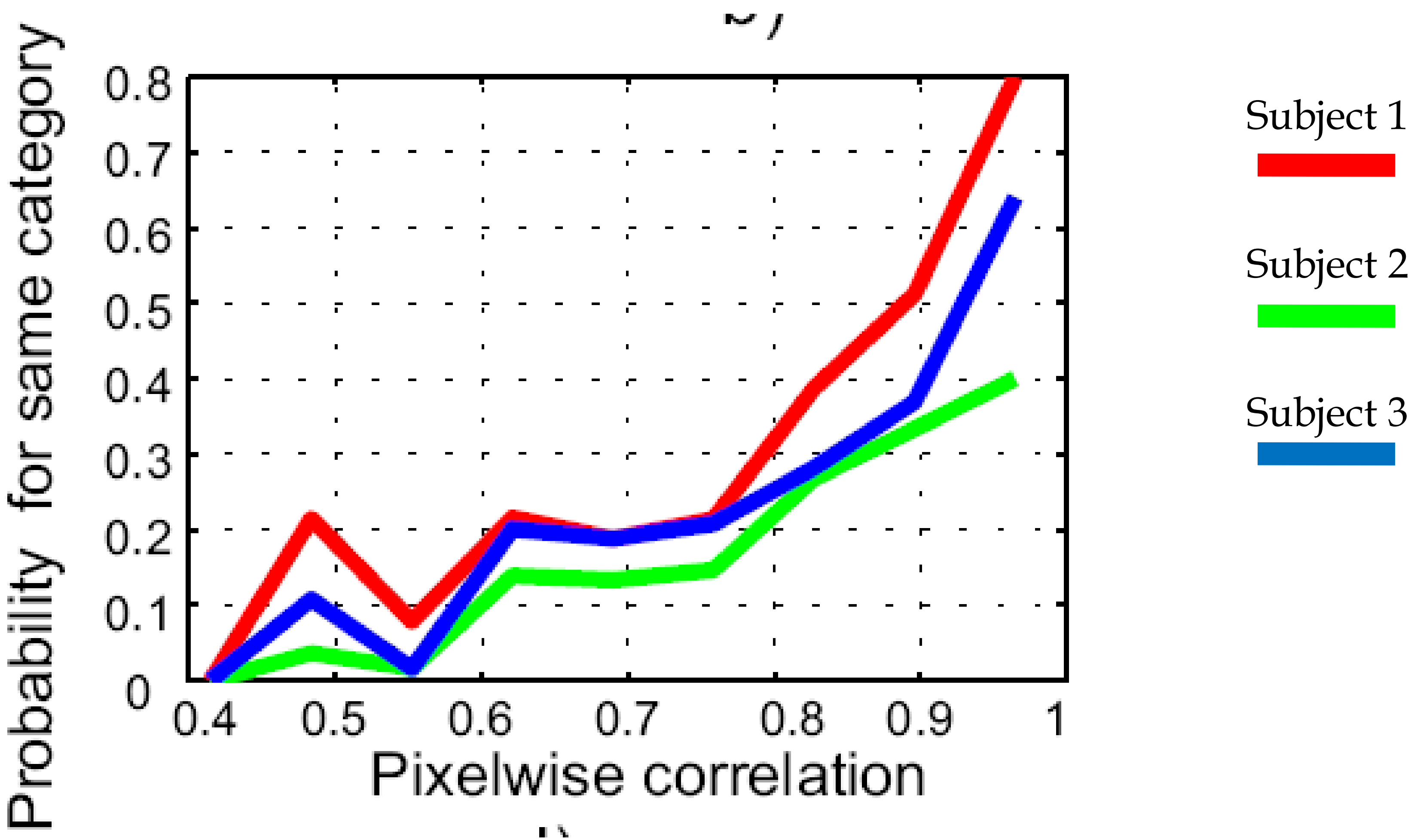
Normalized correlation scores:

How Many Images Are There?



Note: $D_1 = D_{SSD}$

How Does D_{ssd} Relate to Semantic Distance?



Label Assignment

- Distance metrics give set of nearby images
- How to compute label?

- Issues:
  - Labeling noise
  - Keywords can be very specific
    - e.g. yellowfin tuna

Wordnet – a Lexical Dictionary

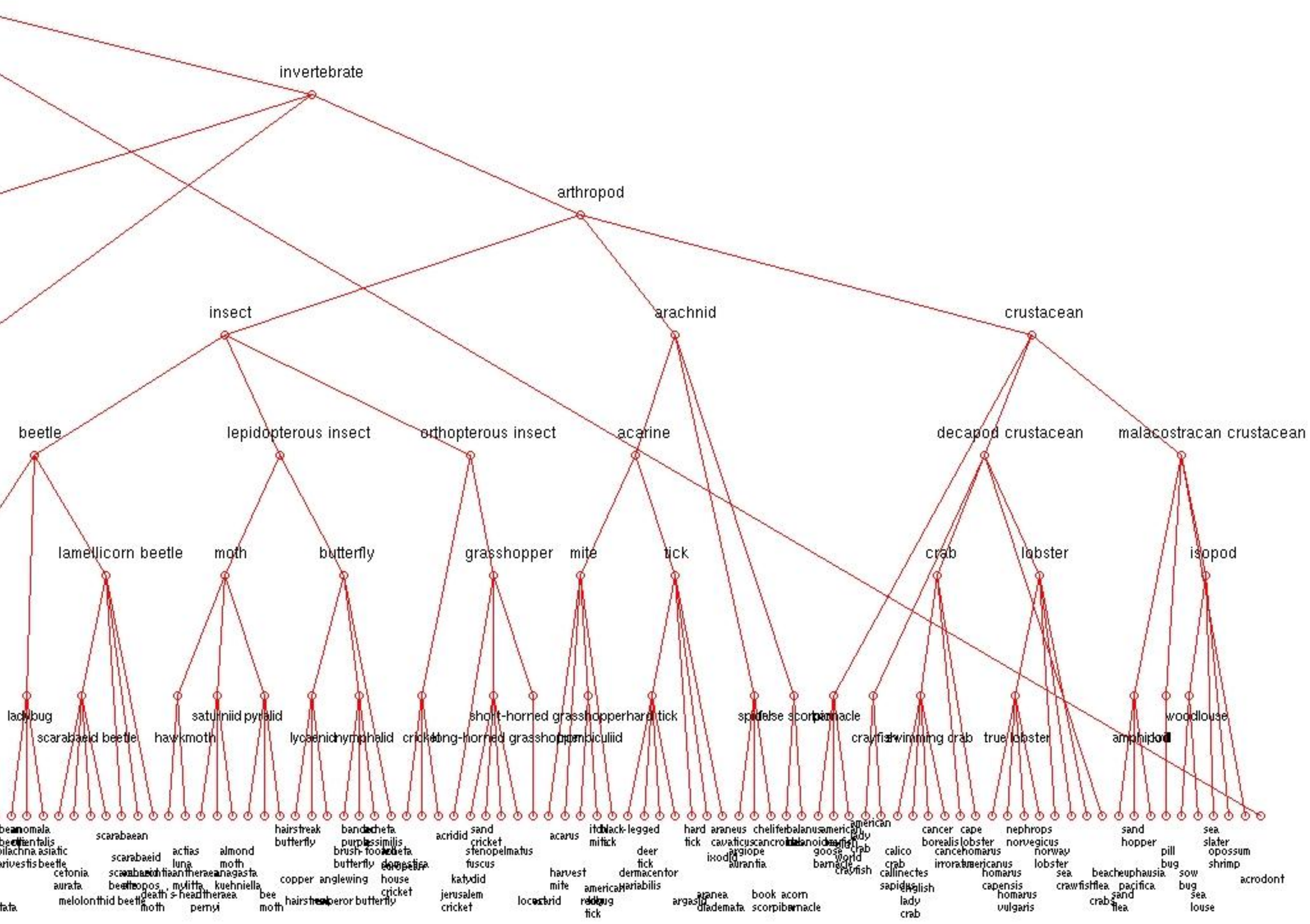
http://wordnet.princeton.edu/

```
Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun aardvark

Sense 1
aardvark, ant bear, anteater, Orycteropus afer
       => placental, placental mammal, eutherian, eutherian mammal
           => mammal
               => vertebrate, craniate
                   => chordate
                       => animal, animate being, beast, brute, creature
                          => organism, being
                               => living thing, animate thing
                                   => object, physical object
                                       => entity
```

Wordnet Hierarchy

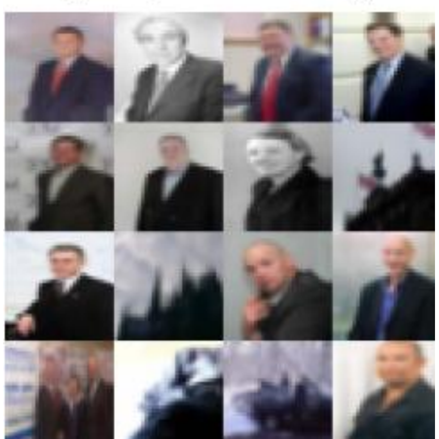
```
Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun aardvark

Sense 1
aardvark, ant bear, anteater, Orycteropus afer
       => placental, placental mammal, eutherian, eutherian mammal
           => mammal
               => vertebrate, craniate
                   => chordate
                       => animal, animate being, beast, brute, creature
                          => organism, being
                               => living thing, animate thing
                                   => object, physical object
                                       => entity
```

- Convert graph structure into tree by taking most common meaning

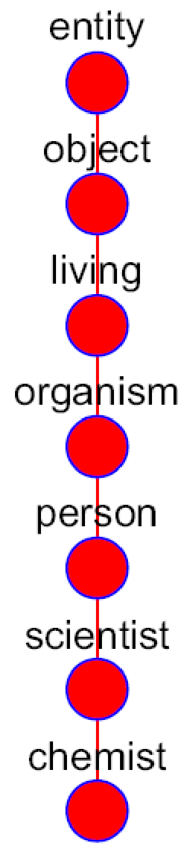

invertebrate
arthropod
insect
arachnid
crustacean
beetle
lepidopterous insect
orthopterous insect
acarine
decapod crustacean
malacostracan crustacean
lamellicorn beetle
moth
butterfly
grasshopper
mite
tick
crab
lobster
isopod
ladybug
scarabaeid beetle
hawkmoth
saturniid
pyralid
lycaenid
nymphalid
cricket
short-horned grasshopper
hard tick
woodlouse
amphipod
crayfish
swimming crab
true lobster
scarabaean
actias luna
almond moth
hairstreak butterfly
acridid
sand cricket
stenopelmatus fuscus
acarus
hard tick
deer tick
dermacentor variabilis
harvest mite
cancer borealis
cape lobster
nephrops norvegicus
norway lobster
sand hopper
sea slater
opossum shrimp
pill bug
sow bug
sea louse
acrodont
calico crab
callinectes sapidus
english lady crab
homarus capensis
homarus vulgaris
euphausia pacifica
sand flea
copper
anglewing
house cricket
katydid
jerusalem cricket
book scorpion
acorn barnacle
aranea diademata
argiope aurantia
goose barnacle
cetonia aurata
melolonthid beetle
bee moth
kuehniella

Wordnet Voting Scheme

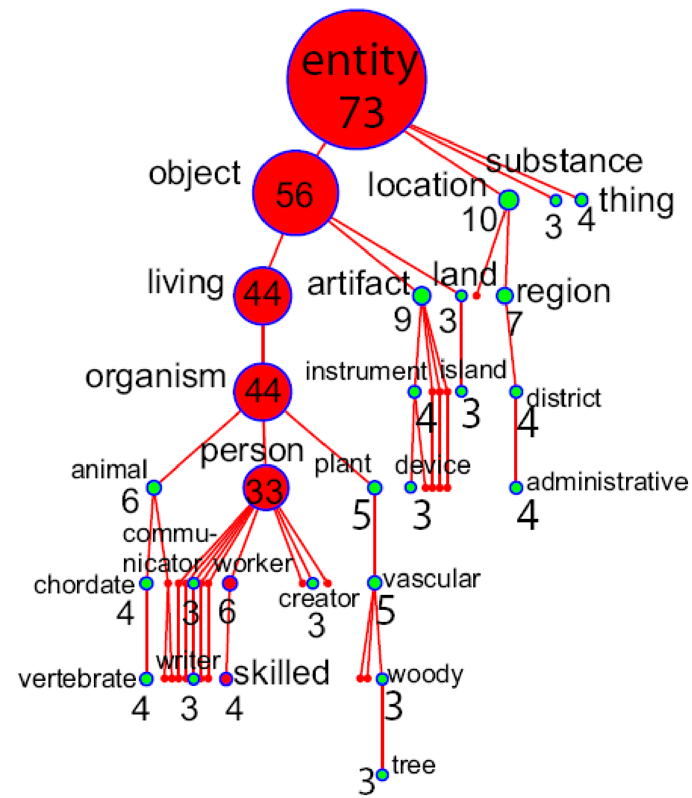
a) Input image



b) Neighbors



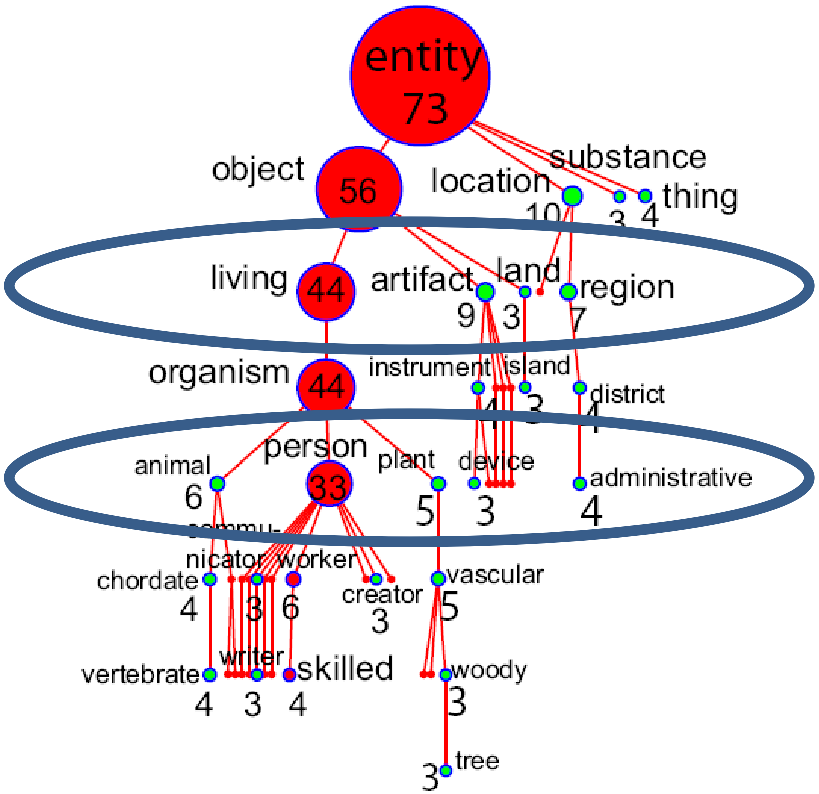
Ground truth



d) Wordnet voted branches

One image – one vote

Classification at Multiple Semantic Levels



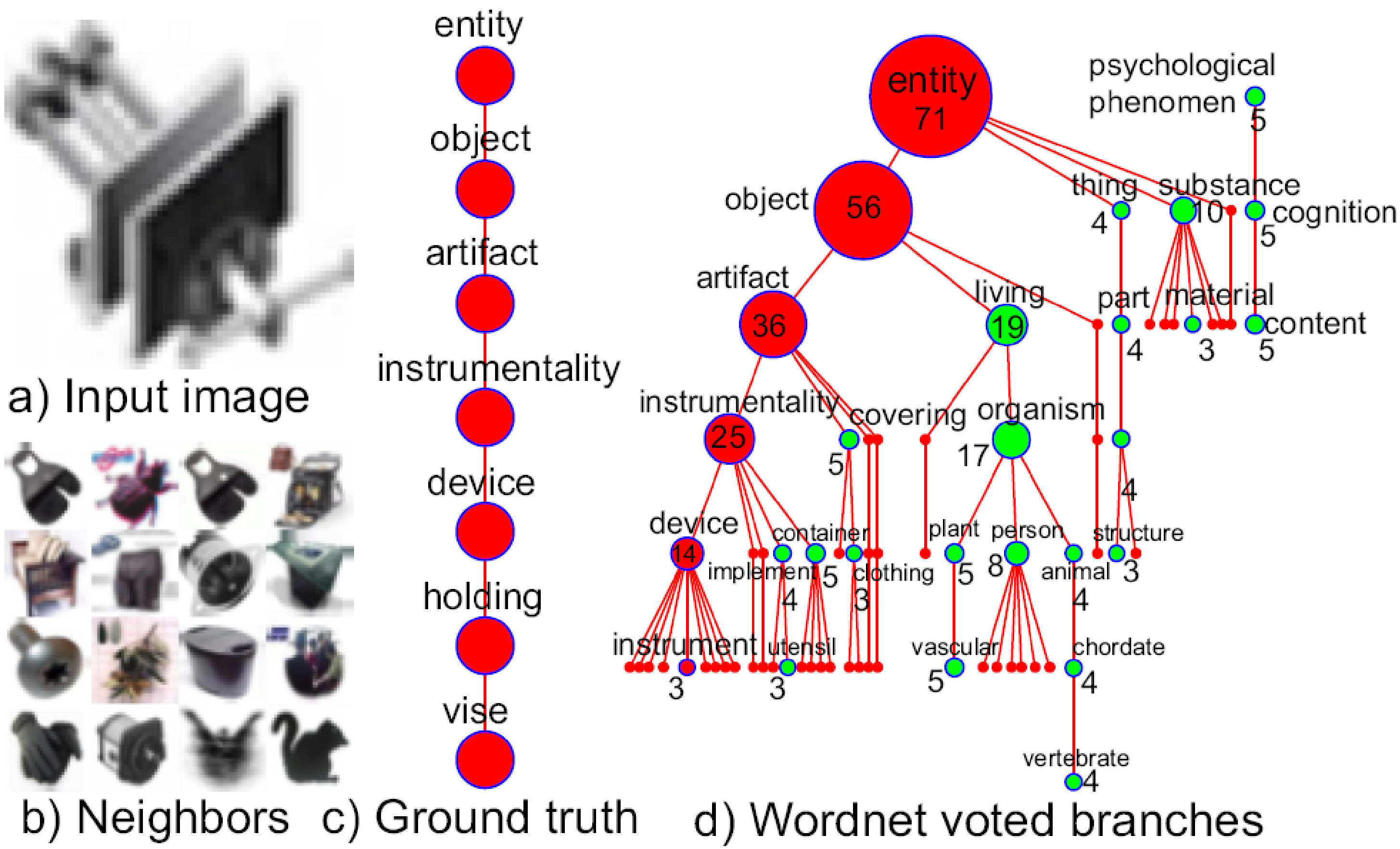
d) Wordnet voted branches

Votes:

Living	44
Artifact	9
Land	3
Region	7
Others	10

Animal	6
Person	33
Plant	5
Device	3
Administrative	4
Others	22

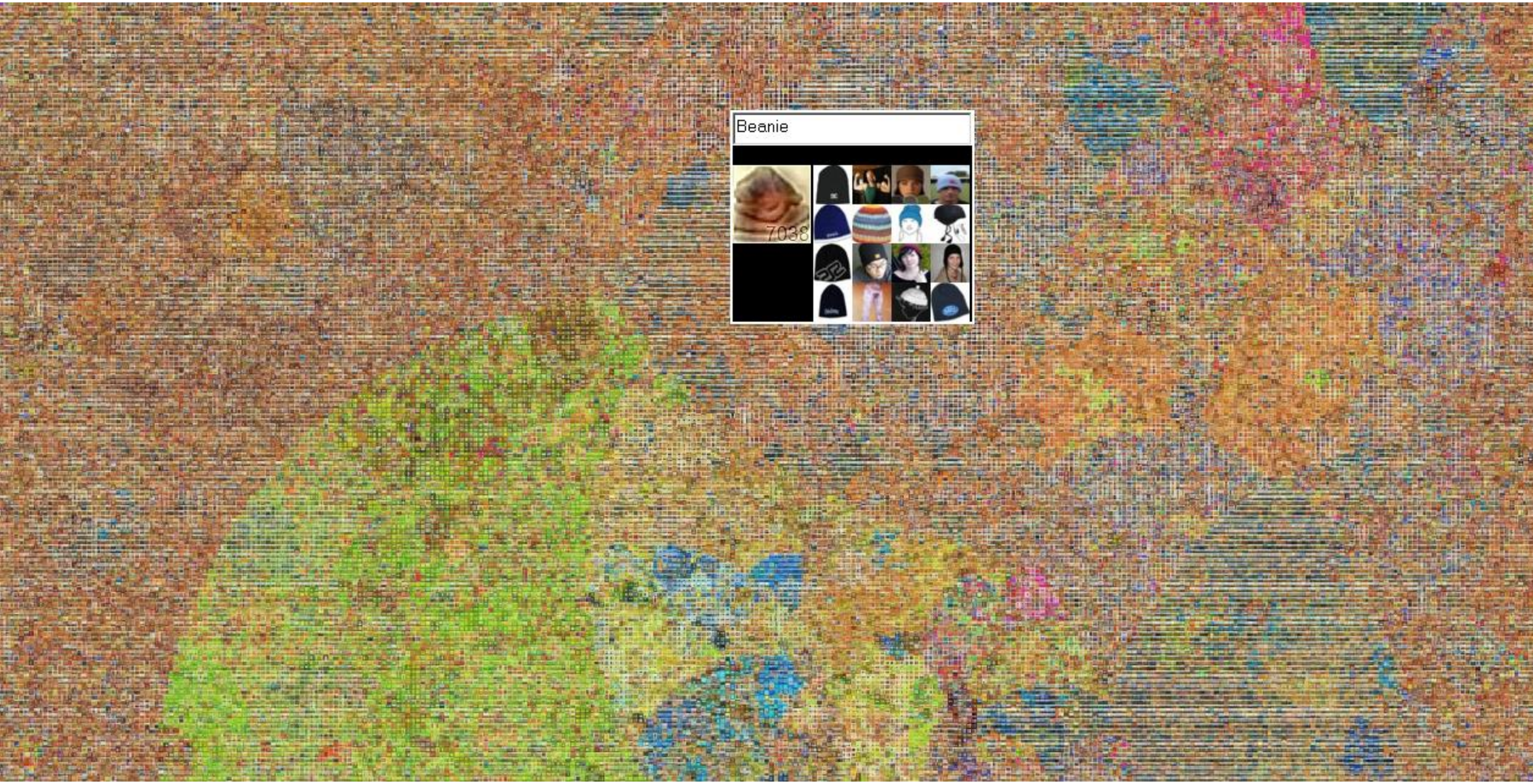
Wordnet Voting Scheme



Wordnet Voting

- Overcomes differences in level of semantic labeling:
  - e.g. "person" & "sir arthur conan doyle"
- Totally incorrect labels form hopefully uniform background noise
- Assumes semantic and visual consistency are closely related

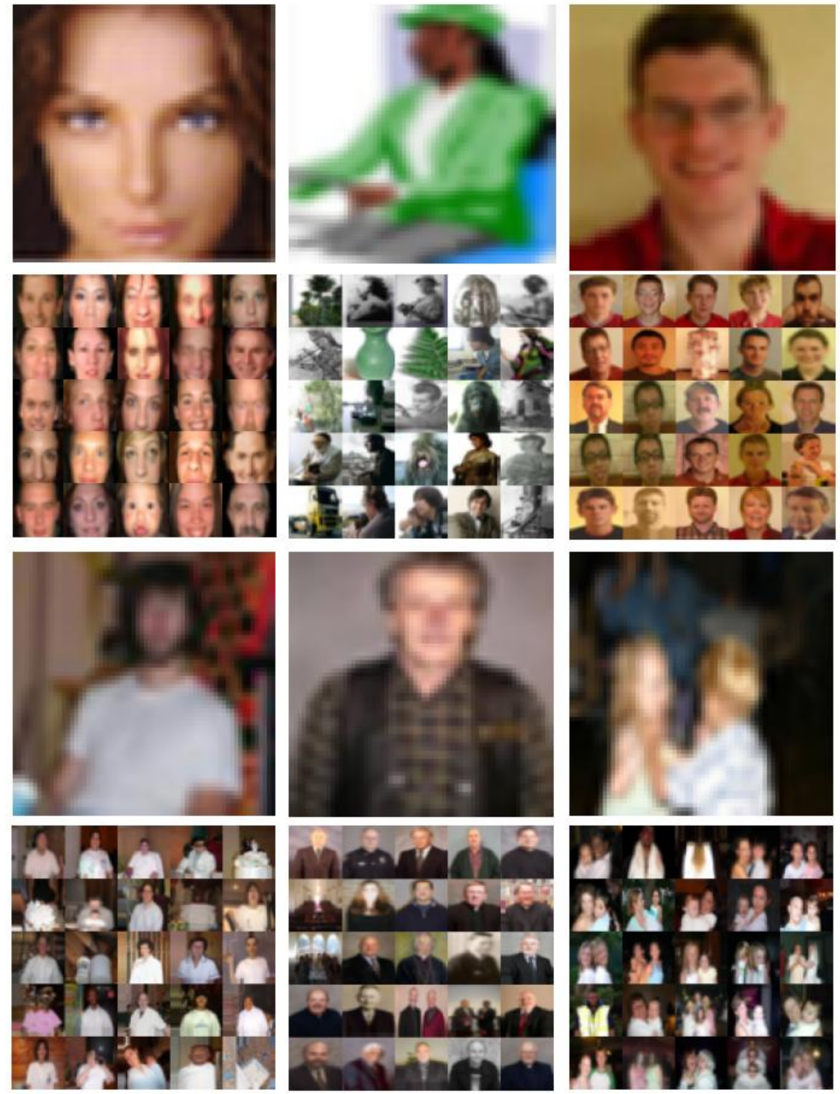
Semantic vs Visual Hierarchy



Recognition Experiments

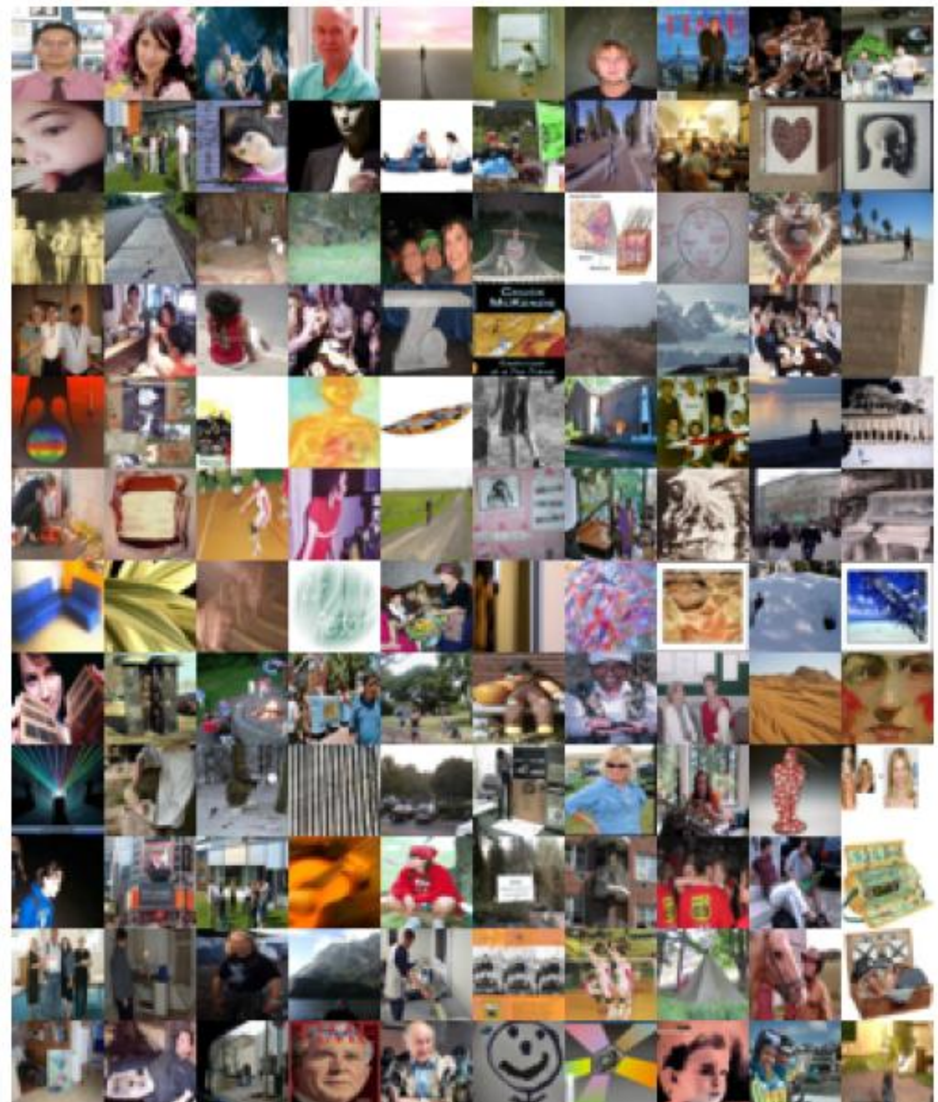
Person Recognition

- 23% of all images in dataset contain people
- Wide range of poses: not just frontal faces



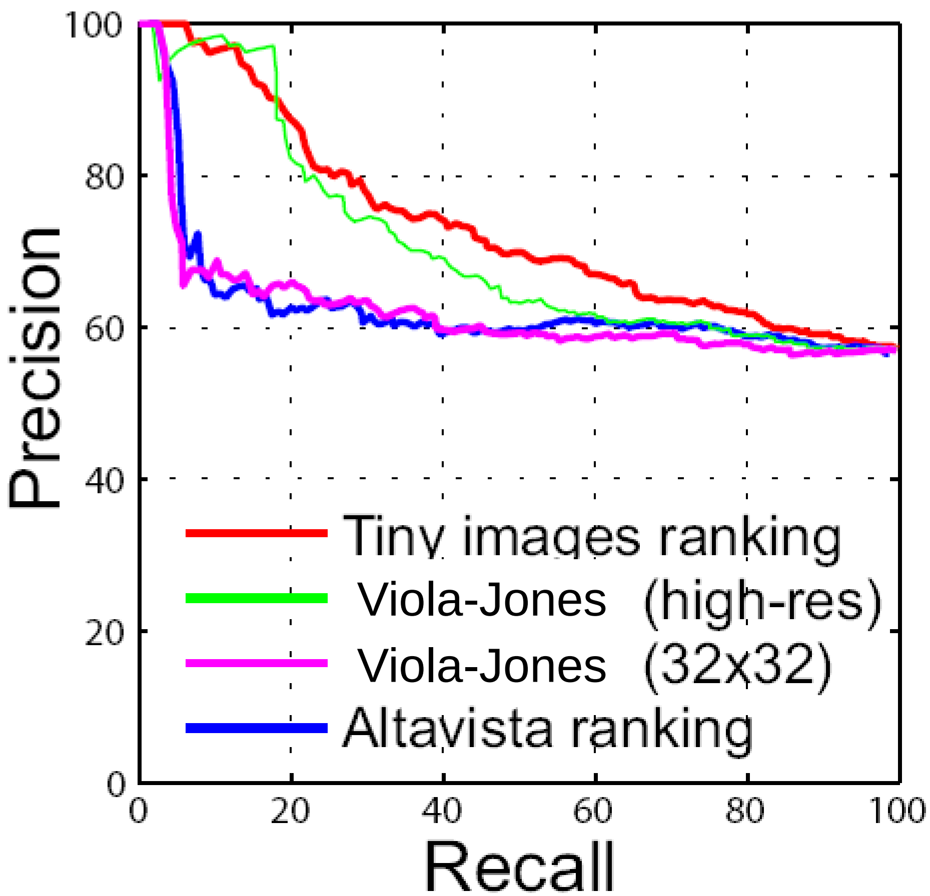
Person Recognition – Test Set

- 1016 images from Altavista using "person" query
- High res and 32x32 available
- Disjoint from 79 million tiny images



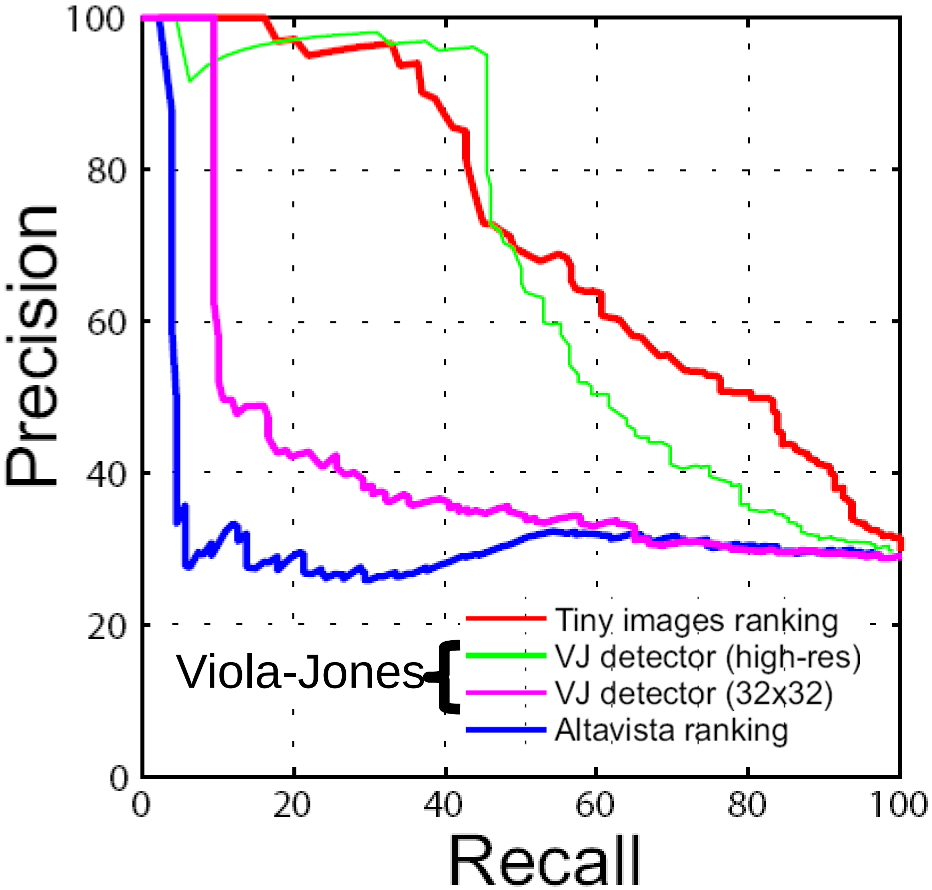
Person Recognition

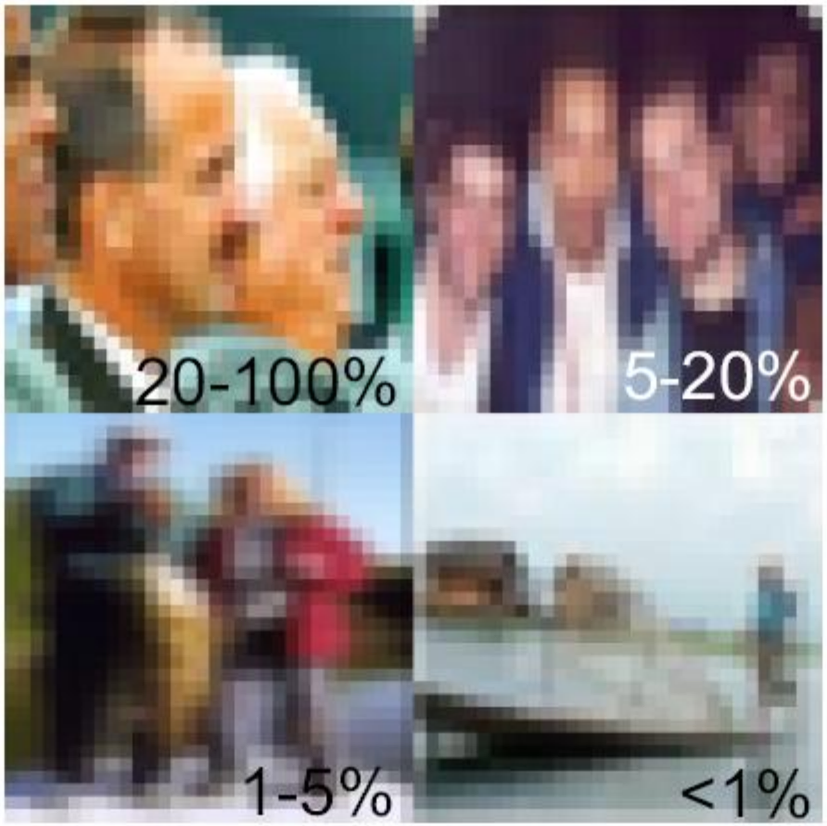
- Task: person in image or not?



Person Recognition

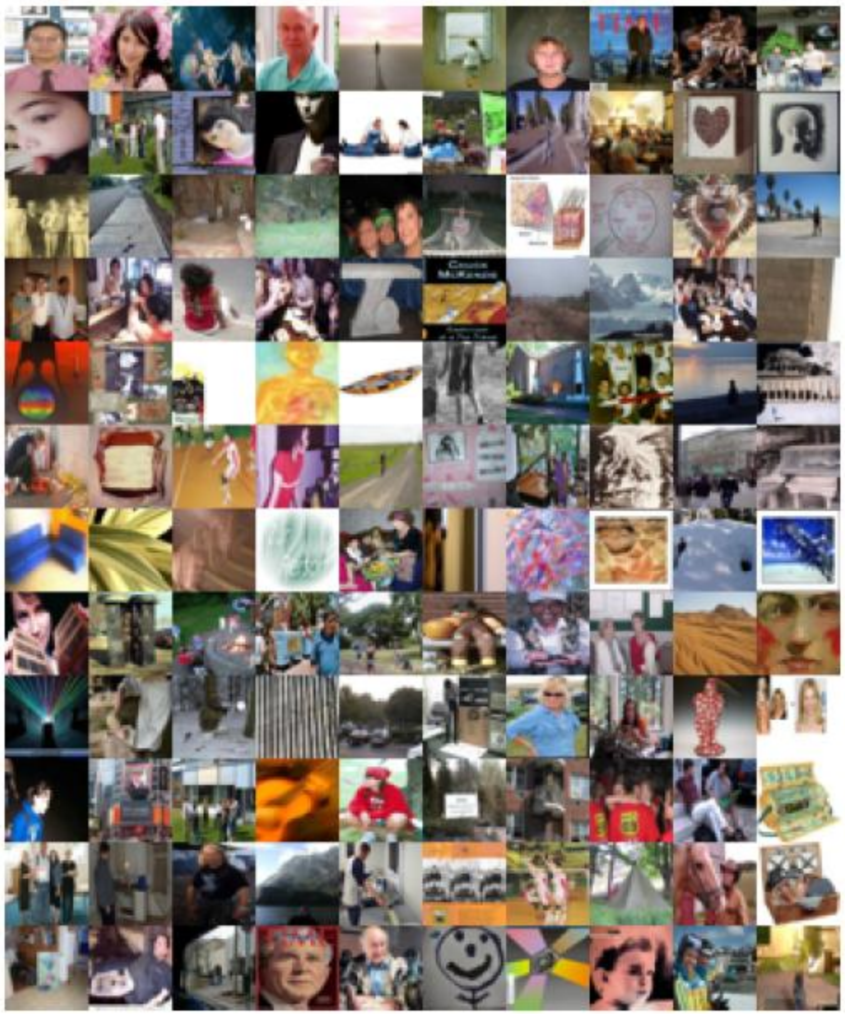
- Subset where face >20% of image





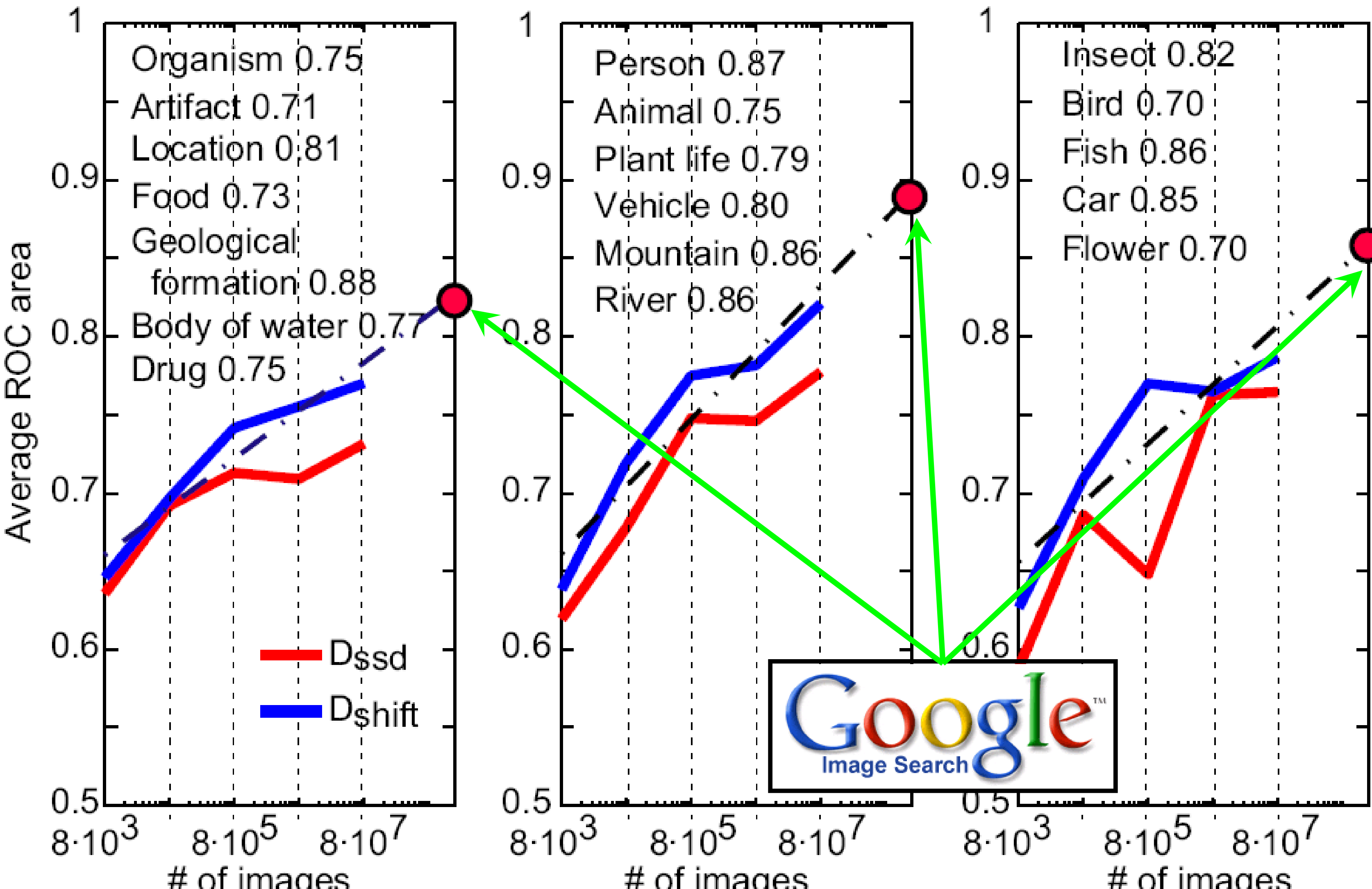
Re-ranked Altavista Images

Original



Re-ranked

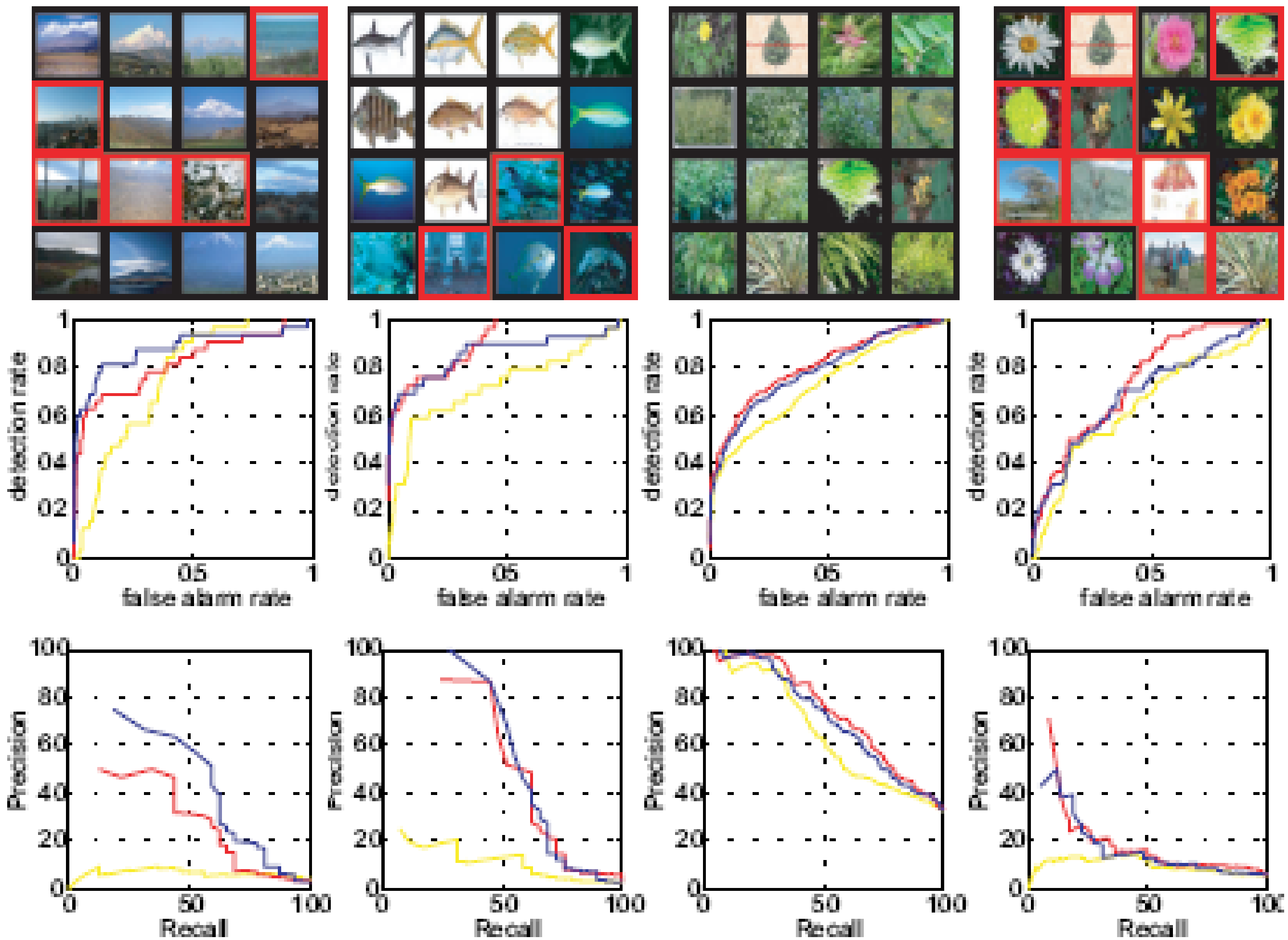
Object Classification



Object Classification

images: 7,900 790,000 79,000,000

Geological formation (32) | Fish (29) | Plant life (335) | Flower (58)



Other Applications

Automatic Colorization

Grayscale input
High resolution

Automatic Colorization

Grayscale input
High resolution

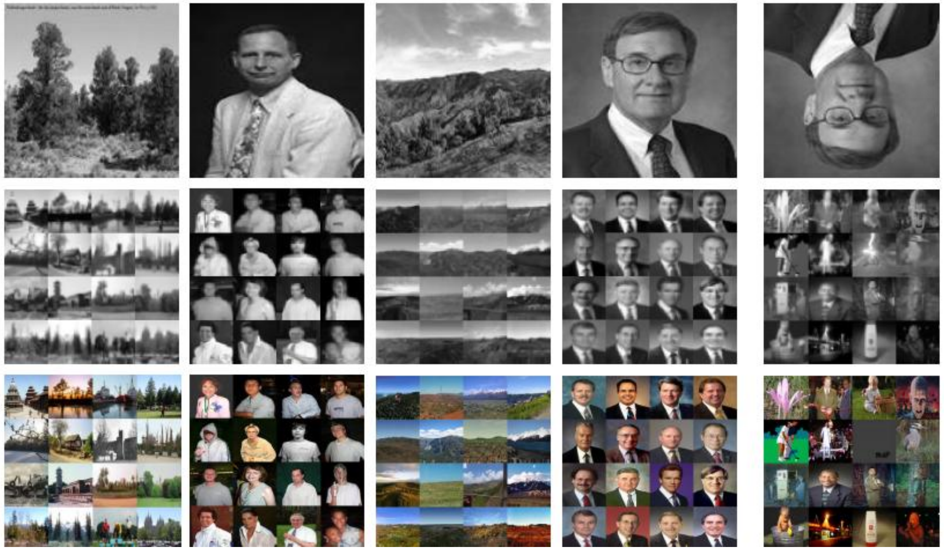
Grayscale
32x32 siblings

Automatic Colorization

Grayscale input
High resolution

Grayscale
32x32 siblings

Color siblings
high resolution



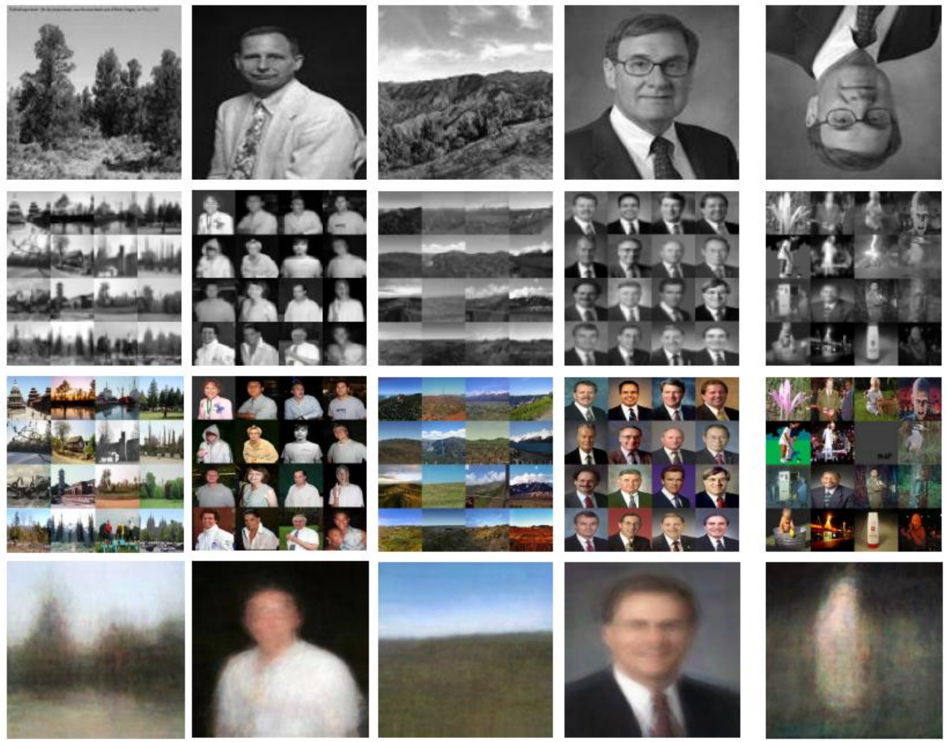
Automatic Colorization

Grayscale input
High resolution

Grayscale
32x32 siblings

Color siblings
high resolution

Average of
color siblings



Automatic Colorization

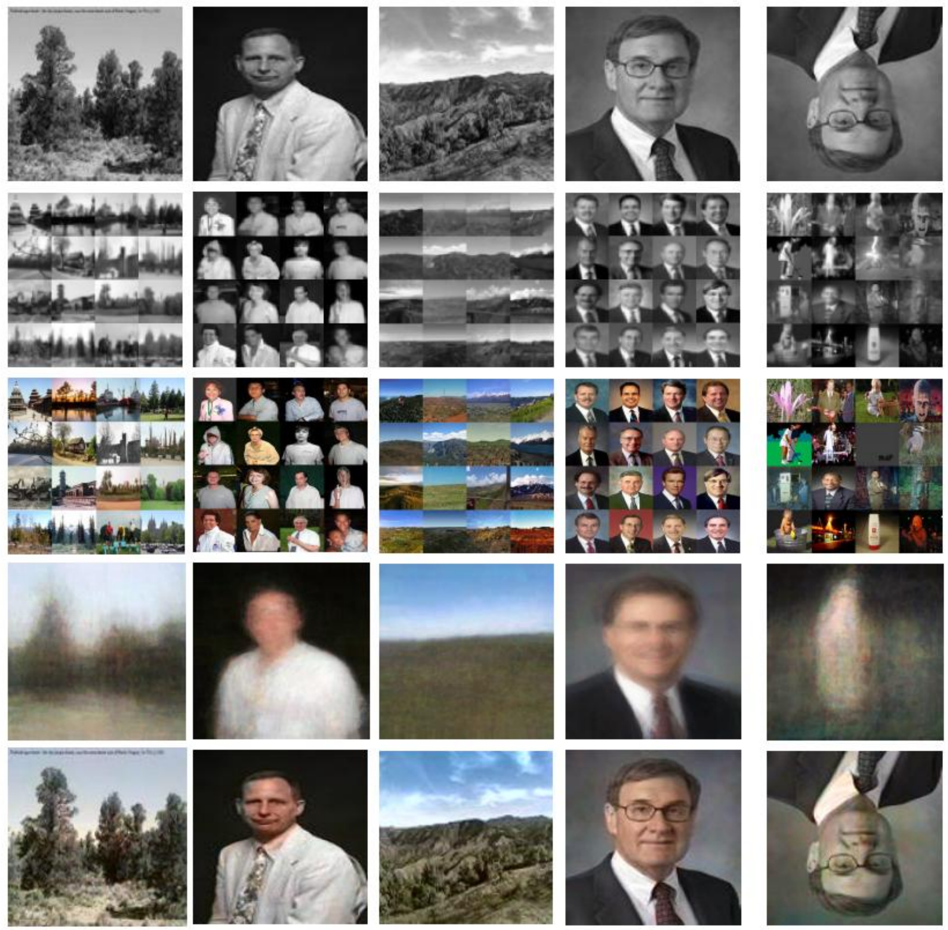
Grayscale input
High resolution

Grayscale
32x32 siblings

Color siblings
high resolution

Average of
color siblings

Colorization of input
using average



Automatic Colorization

Grayscale input
High resolution

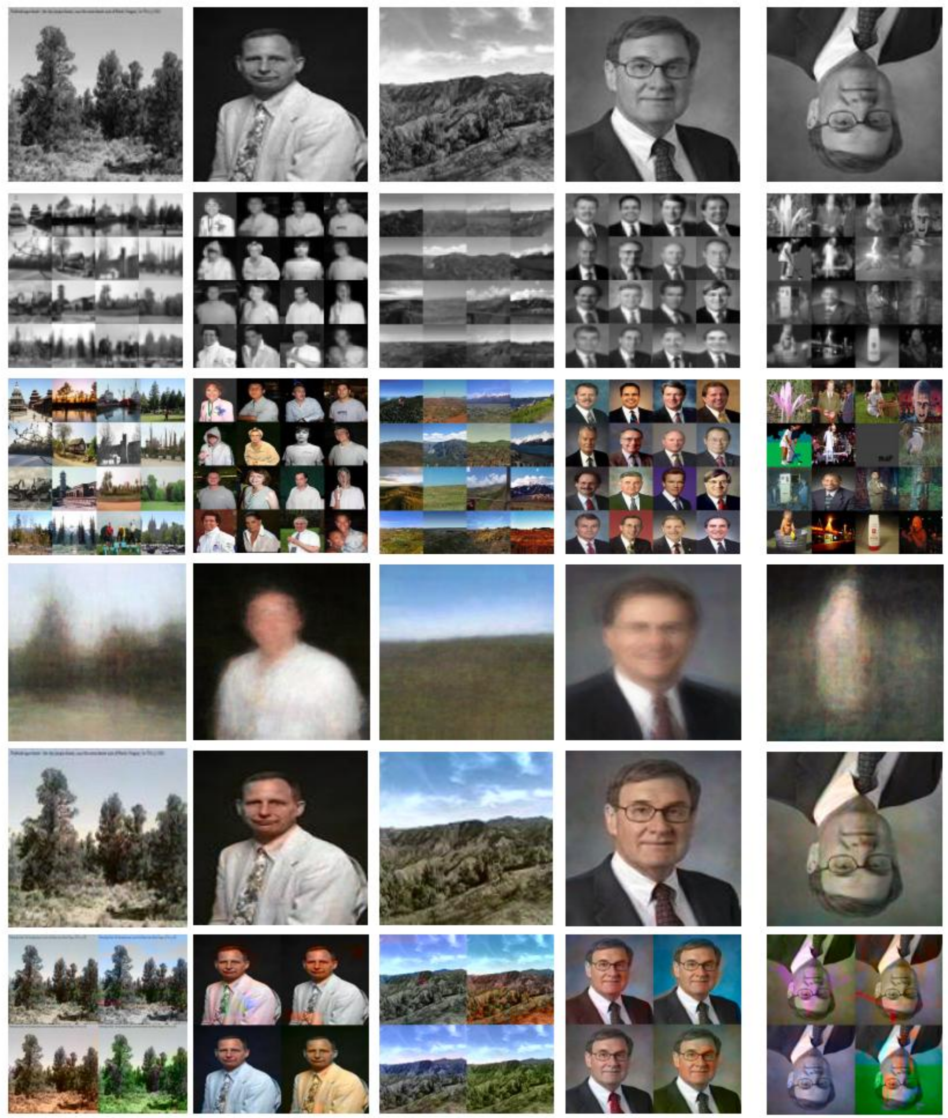
Grayscale
32x32 siblings

Color siblings
high resolution

Average of
color siblings

Colorization of input
using average

Colorization of input
using specific siblings



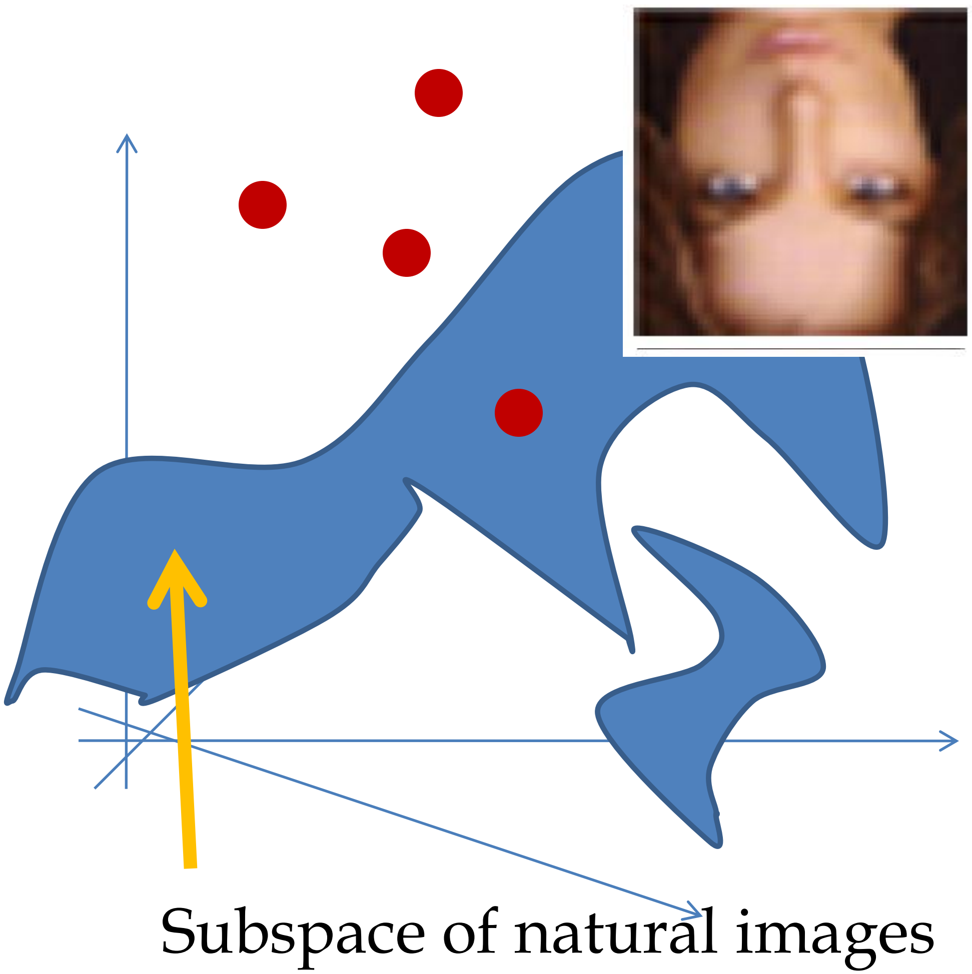
Automatic Colorization Result

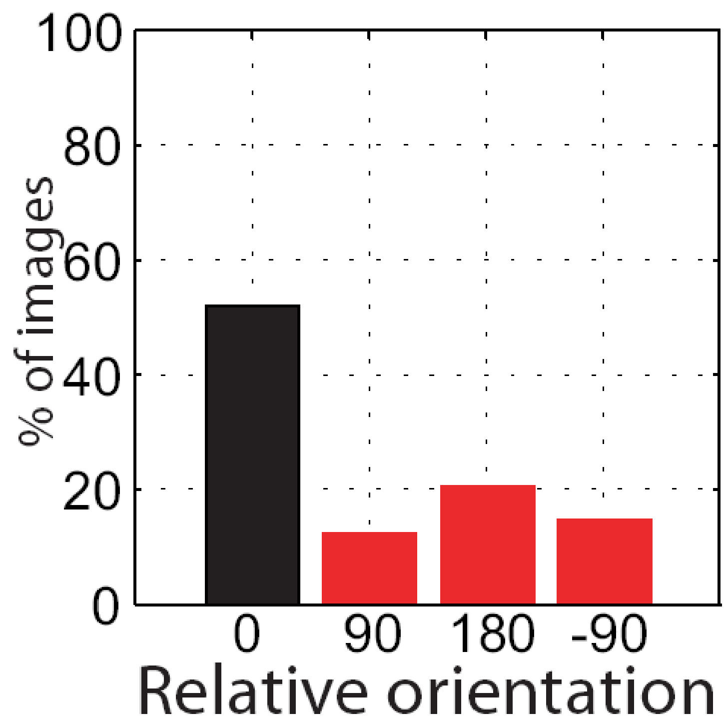
Grayscale input High resolution

Colorization of input using average

Automatic Orientation

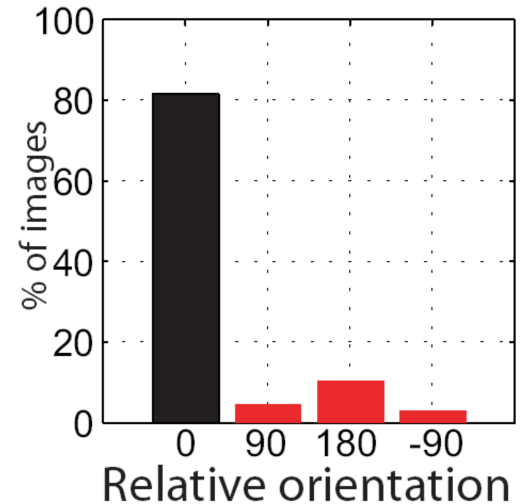
- Look at mean distance to neighbors





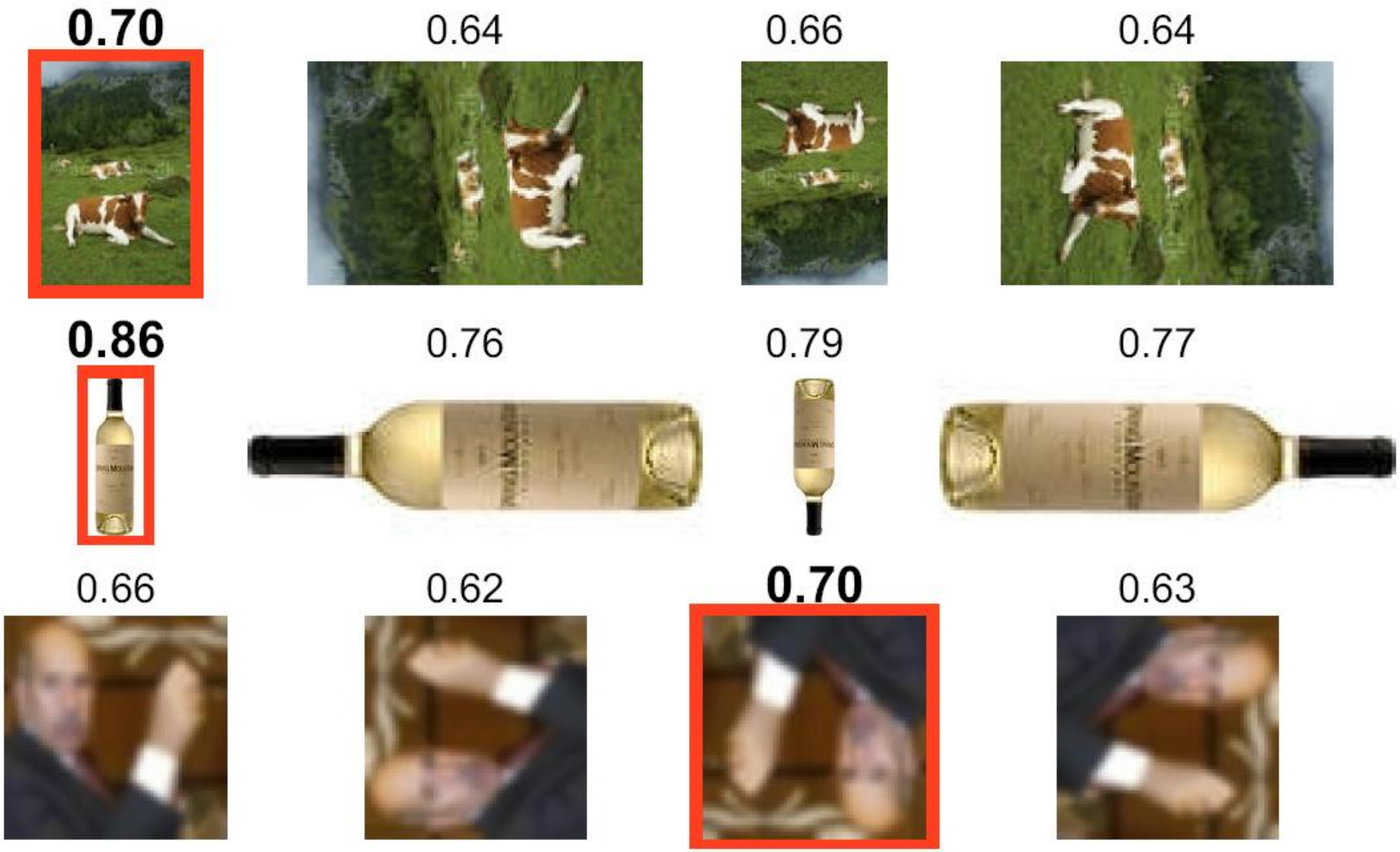
Automatic Orientation

- Many images have ambiguous orientation
- Look at top 25% by confidence:



- Examples of high and low confidence images:

Automatic Orientation Examples



Related Work

- Hayes & Efros, Scene Completion using Millions of photographs, SIGGRAPH 2007.
- Nister & Stewenius. Scalable recognition with a vocabulary tree, CVPR 2006.
- Hoogs & Collins. Object boundary detection in images using a semantic ontology. In *AAAI, 2006.*
- Barnard et al., Matching words and pictures. JMLR, 2003.
- Shakhnarovich et al. Fast pose estimation with parameter sensitive hashing, ICCV 2003

Conclusions

Few Data
Complex Model

Huge amounts of Data
No Model

- Can get good results simple algorithms & lots of data